

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 04:43:36 ; Search time 46.76 Seconds  
(without alignments)  
1308.014 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 1 cagcggtygtgcaggaagc.....gtctattttttttatttg 249

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_6/prodata/1/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/prodata/1/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/prodata/1/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/prodata/1/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/prodata/1/ina/PCTUS.COMB.seq.\*
- 6: /cgn2\_6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	36	14.5	2529	US-09-051-969A-5
2	35.2	14.1	619	US-09-328-111-579
3	31.4	12.6	579	US-09-146-950-3
4	31.4	12.6	591	US-09-146-950-19
5	31.4	12.6	1596	US-09-146-950-17
6	31.4	12.6	1724	US-08-509-024-1
7	31.4	12.6	1724	US-09-333-279-1
8	31.4	12.6	1929	US-09-146-950-1
9	31.4	12.6	4622	US-08-509-024-6
10	31.4	12.6	4622	US-09-333-279-6
11	30.4	12.2	2664	US-08-539-304A-5
12	30.4	12.2	2663	US-08-136-743B-3
13	30.4	12.2	3252	US-09-118-442-1
14	30.4	12.2	3252	US-09-677-064-1
15	29.8	12.0	1724	PCT-US96-12374-1
16	29.6	11.9	860	US-07-847-010-18
17	29.6	11.9	2658	US-08-773-608A-1
18	29.4	11.8	524	PCT-US95-08295-22
19	29.4	11.8	4875	US-08-460-739-1
20	29.2	11.7	19124	US-08-487-826B-13
21	29	11.6	279	US-08-686-878A-50
22	29	11.6	279	US-08-721-489-4
23	28.8	11.6	684	US-08-577-463A-1
24	28.8	11.6	2946	US-09-175-928-3
25	28.4	11.4	2622	US-08-766-014-23
26	28.4	11.4	2902	US-08-714-918-95
27	28.4	11.4	2902	US-09-265-315-95

C 28	28.4	11.4	2902	4	US-09-265-315-95	Sequence 95, Appl
C 29	28.4	11.4	2902	4	US-09-266-417-95	Sequence 95, Appl
C 30	28.4	11.4	5300	1	US-08-766-014-1	Sequence 1, Appl
C 31	28.2	11.3	618	4	US-08-961-810-17	Sequence 17, Appl
C 32	28.2	11.3	618	4	US-08-352-902D-17	Sequence 17, Appl
C 33	28.2	11.3	2075	1	US-08-238-163-3	Sequence 3, Appl
C 34	28	11.2	392	1	US-08-468-421-8	Sequence 8, Appl
C 35	28	11.2	392	1	US-08-250-975-8	Sequence 8, Appl
C 36	28	11.2	392	2	US-08-605-002A-8	Sequence 8, Appl
C 37	28	11.2	392	2	US-08-950-449A-8	Sequence 8, Appl
C 38	28	11.2	392	5	PCT-US94-10529-8	Sequence 8, Appl
C 39	28	11.2	2612	4	US-09-105-390-7	Sequence 7, Appl
C 40	28	11.2	6124	4	US-08-213-419B-3	Sequence 3, Appl
C 41	27.8	11.2	886	1	US-08-469-427A-1	Sequence 1, Appl
C 42	27.8	11.2	886	2	US-08-609-443B-1	Sequence 1, Appl
C 43	27.8	11.2	886	2	US-08-569-063C-1	Sequence 1, Appl
C 44	27.8	11.2	1002	4	US-08-960-780-43	Sequence 43, Appl
C 45	27.8	11.2	1002	4	US-09-073-898-43	Sequence 43, Appl

# ALIGNMENTS

```

RESULT 1
US-09-051-969A-5
Sequence 5, Application US/09051969A
Patent No. 6063598
GENERAL INFORMATION:
APPLICANT: ENENKEL, BARBARA
APPLICANT: GANNON, FRANK
APPLICANT: BERGMANN, KLAUS
APPLICANT: NOE, WOLFGANG
TITLE OF INVENTION: INTENSIVE HOMOLOGOUS PROMOTER OBTAINED
TITLE OF INVENTION: FROM HAMSTERS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051, 969A
FILING DATE: 1998-09-30
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FLESHNER, RAZ E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 0652.1690000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2529 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2418..2465
US-09-051-969A-5
Query Match 14.5%; Score 36; DB 3; Length 2529;
Best Local Similarity 50.9%; Pred. No. 0.042;

```



```

: Patent No. 6287808
:
: GENERAL INFORMATION:
:
: APPLICANT: Busfield, Samantha J.
:
: TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
:
: TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
:
: FILE REFERENCE: 09404/057001
:
: CURRENT APPLICATION NUMBER: US/09/146,950A
:
: CURRENT FILING DATE: 1998-09-03
:
: NUMBER OF SEQ ID NOS: 25
:
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ ID NO 17
:
: LENGTH: 1596
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
US-09-146-950-17

```

	Query Match	Best Local Similarity	Score 31.4%	DB 4	Length 1596
	Matches 74	Conservative 0	Mismatches 71	Indels 0	Gaps 0
QY	53	gggtctcggggatctcggctggtcccatcttccttcattgtgtctgaacatcctgtatgt	112		
Db	587	gggtgtccttgactctcggtggtccctcttctgcacaccttgagccggggctgagtggtcgt	528		
QY	113	aaacatcgtcgtgggtgtgctaaagtgcctgtgtaatcccgatgtggaanaagctcggaggtga	172		
Db	527	AAGCCGGCAGCCGGCGGCGAGTGGTCCCTCCTCGACGATGAGAGAGTGGCTGGCTGC	468		
QY	173	aagctcagcatatcattatattc	197		
Db	467	AGCCACACAGCGCTTCTCTGTCT	443		

```

US-08-509-024-1/c
US-08-509-024-1
6
Sequence 1, Application US/08509024B
Patent No. 6291207
GENERAL INFORMATION:
APPLICANT: SPEAR, Patricia G.
APPLICANT: MONTGOMERY, Rebecca I.
TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
FILE REFERENCE: 0290-1
CURRENT APPLICATION NUMBER: US/08/509,024B
CURRENT FILING DATE: 1995-07-25
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1724
TYPE: DNA
ORGANISM: Homo sapiens
US-08-509-024-1

```

Query Match	Similarity	Score	DB	Length
Best Local	51.0%	Pred. 1.1		
Matches	74; Conservative	0; Mismatches	71; Indels	0; Gaps
QY	53	gggtctcgagatctcgtgcgtcccatcttccttcatgttctgaacatcctgtattgt	112	
DB	774	gggtgctccgactctggtgctcctcttctgcacccctctgccccggggctggagctggcgt	715	
QY	113	aaacatggtcgtggtgtctaaagtgcctgtgaatcccgatgtgaaaaagctgaggtga	172	
DB	714	aagcggcgacacggcgccagtggtccctccctccgacgatgacgaagtggcctcgccctgc	655	
QY	173	aagctcagcatatcatgtatttact	197	
DB	654	agccacacacggcggtctctgtctt	630	

RESULT 7  
US-09-333-279-1/c

```

Sequence 1 Application US/093333279
Patent No. 6303336
GENERAL INFORMATION:
APPLICANT: SPEAR, Patricia G.
APPLICANT: MONTGOMERY, Rebecca I.
TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
FILE REFERENCE: 0290-1
CURRENT APPLICATION NUMBER: US/09/333,279
CURRENT FILING DATE: 1999-06-15
NUMBER OF SEQ. ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
Seq ID NO 1
LENGTH: 1724
TYPE: DNA
ORGANISM: Homo sapiens
US-09-333-279-1

```

Query Match	12.6%	Score 31.4	DB 4	Length 1724
Best Local Similarity	51.0%	Pred. NO. 1.1	71	Indels 0
Matches 74	Conservative	0	Mismatches	Gaps 0
QY 53	ggatcgcggagatcgcgtgcgtcccaatcttcattcgtgtctgaacatccgtgaatgt	112		
DB 774	gggtttctctgactctcgggtgccttcctcttgacaccttgagcccgcggtgagagtgacct	715		
QY 113	aaacacatgcgtctgggtgtcctaagtcctcgttgaatcccgatgttggaaaagctcgaagtga	172		
DB 714	aagccgcgcacagcggcgacgagtggtcccgctctgcagcagtcgacaagtgccctggcgctgc	655		
QY 173	aagctaaagcatcacatcgtfaattact	197		
DB 654	agccacacacagcgcttctctgtctct	630		

```

RESULT      8
US-09-146-950-1/c
; Sequence 1, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Bustfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1929
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (297)...(875)
US-09-146-950-1

```

Query Match	12.6%	Score 31.4	DB 4	length 1929
Best Local Similarity	51.0%	Pred. No. 1.1	Mismatches	71
Matches	74	Conservative	0	Indels 0
Qy	53	gggtctcggagatcctcggatcgcctccatctccatctatgtctcgaacatcctgtatgt	112	
Db	777	gggtgtcccgactctccgtccggccctccctcttgacacctcgaccggcggagtgccgt	718	
Qy	113	aaacatgagctggggtgctaaagtcctgtgaatcccgatgtagaaagctggagtgta	172	
Db	717	aagccggcgacacggcgcgatgggtcccgccgcgagcagatgcagaagtgccctggcctgc	658	
Qy	173	aagctcagcatcatctgtatttact	197	
Db	657	agccacacacggcggtctctgtcct	633	



```

; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The University of Pennsylvania
; STREET: Suite 330
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19104-3246
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 KB
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,743B
; FILING DATE: 10/14/93
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 3957-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5459063e
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2663 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-136-743B-3

```

```

Query Match      12.2%; Score 30.4; DB 1; Length 2663;
Best Local Similarity 63.9%; Pred. No. 2.8;
Matches 46; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 176 ctacgacatcgtatattacttaaacagaaagacatgatagtctcat 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2301 CTTATTAACAAATGACATACATATATAATATATATATATATATAT 2242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 236 tttttttttt 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2241 ATTTTTTTTTT 2230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-09-118-442-1/C
; Sequence 1, Application US/09118442B
; Patent No. 6197561
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; APPLICANT: Wang, Xun
; APPLICANT: Bowen, Benjamin A.
; TITLE OF INVENTION: Genes Controlling Phylate Metabolism In
; TITLE OF INVENTION: Plants and Uses Thereof
; FILE REFERENCE: 0706
; CURRENT APPLICATION NUMBER: US/09/118,442B
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/055,446
; EARLIER FILING DATE: 1997-08-11
; EARLIER APPLICATION NUMBER: 60/055,526
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: 60/053,944
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1

```

```

; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (258)...(2666)
; NAME/KEY: misc_feature
; LOCATION: (1)...(3252)
; OTHER INFORMATION: n = A,T,C or G
;
; US-09-118-442-1

```

```

Query Match      12.2%; Score 30.4; DB 4; Length 3252;
Best Local Similarity 50.0%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 92 gtctgaacatcgtatattgttaaacatgctggggtgctaaagtgcgtgaatccga 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1862 GTTCAAGCAGCTTAATAATATAGAGATCTTGAGGACACACTCTGTTAGAGAGAGT 1803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 tttggaagaaagctggagtggaagctcagcatcacatgattactttaaacagaaan 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1802 TGGTCTTAATGATGAACGATTTGGCTCATCAAGATTGTGAAGCTCACTGAAGACTCTGA 1743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 212 aaagacatgatagtatgtcattttttt 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1742 TAATAGCTGCCATTTTTCATTTTCTTT 1711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
US-09-677-064-1/C
; Sequence 1, Application US/09677064
; Patent No. 6291224
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; TITLE OF INVENTION: Genes Controlling Phylate Metabolism In
; TITLE OF INVENTION: Plants and Uses Thereof
; FILE REFERENCE: 0706D
; CURRENT APPLICATION NUMBER: US/09/677,064
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (258)...(2666)
; NAME/KEY: misc_feature
; LOCATION: (1)...(3252)
; OTHER INFORMATION: n = A,T,C or G
;
; US-09-677-064-1

```

```

Query Match      12.2%; Score 30.4; DB 4; Length 3252;
Best Local Similarity 50.0%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 92 gtctgaacatcgtatattgttaaacatgctggggtgctaaagtgcgtgaatccga 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1862 GTTCAAGCAGCTTAATAATATAGAGATCTTGAGGACACACTCTGTTAGAGAGAGT 1807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 04:33:10 ; Search time 1809.29 Seconds  
(without alignments)  
1857.492 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 249  
Sequence: 1 cagcggtgtgtgcaggaagc.....gtctattttttttatcgtg 249

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_estl:\*  
10: gb\_estc2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_hiv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	249	100.0	536	9 A1079538	A1079538 oz04e06.x
C 2	249	100.0	536	9 A1497808	A1497808 tm89f02.x
C 3	249	100.0	561	9 A1038061	A1038061 ox21b05.x
C 4	247.4	99.4	626	10 BF679962	BF679962 602154666
C 5	242.6	97.4	618	9 BE176440	BE176440 RC3-H7058
C 6	241	96.8	540	9 AW242451	AW242451 xm899a04.x
C 7	237	95.2	486	9 A1333325	A1333325 qd05b01.x
C 8	237	95.2	466	9 A1952086	A1952086 wx46g11.x
C 9	233.2	93.7	545	9 A1129988	A1129988 qc50c03.x
C 10	233.2	93.7	687	9 BE176633	BE176633 RC3-H7058
C 11	215	86.3	463	9 A1356239	A1356239 qy5f11.x
C 12	208	83.5	462	9 A1309934	A1309934 qy71e02.x
C 13	198	79.5	446	9 A1369277	A1369277 qy91b11.x
C 14	187	75.1	436	9 A1192867	A1192867 qe68b05.x
C 15	176	70.7	427	9 AA293275	AA293275 zt38e06.s
C 16	165	66.3	179	9 A1041288	A1041288 DKFZp34M
C 17	159.2	63.9	419	9 A1739028	A1739028 w134d09.x

C 18	146.2	58.7	400	9 A1002564	A1002564 oq90g04.s
C 19	130.2	52.3	384	9 AA909465	AA909465 ol14f04.s
C 20	129	51.8	378	9 A1590920	A1590920 z057c03.s
C 21	90.2	36.2	482	10 BF039164	BF039164 BP250008B
C 22	83.2	33.4	577	10 BF041786	BF041786 BP250016A
C 23	82	32.9	267	10 BE66034	BE66034 149038 MA
C 24	81.6	32.8	621	10 BF040740	BF040740 BP250010B
C 25	79	31.7	248	9 AM659322	AM659322 96381 MAR
C 26	78	31.3	599	10 BF398487	BF398487 UI-R-B52-
C 27	76.6	30.8	221	10 BM481293	BM481293 533247 MA
C 28	76.2	30.6	492	9 AA759442	AA759442 vw64f04.r
C 29	76	30.5	334	9 A1038665	A1038665 ox39f08.s
C 30	76	30.5	481	10 BG376470	BG376470 UI-R-CU0-
C 31	73	29.3	536	10 BM391560	BM391560 UI-R-DX0-
C 32	73	29.3	591	10 BM383627	BM383627 UI-R-DM1-
C 33	72	28.9	456	9 BE105642	BE105642 UI-R-BX0-
C 34	70.4	28.3	206	9 A1713888	A1713888 UI-R-AG1-
C 35	70.4	28.3	284	9 AM253128	AM253128 UI-R-BJ0-
C 36	70.4	28.3	502	9 AM253386	AM253386 UI-R-BJ0-
C 37	70	28.1	328	9 A1038666	A1038666 ox39f09.s
C 38	69	27.7	563	9 A1103427	A1103427 EST212716
C 39	69	27.7	444	9 A1415711	A1415711 mc74a08.x
C 40	63.6	25.5	458	9 BB830241	BB830241 BB830241
C 41	63.2	25.4	432	9 BB749365	BB749365 BB749365
C 42	62	24.9	321	9 BB837582	BB837582 BB837582
C 43	62	24.9	349	9 BF148538	BF148538 uy82e08.x
C 44	62	24.9	388	9 BB836719	BB836719 BB836719
C 45	62	24.9	411	9 AA794533	AA794533 vu68f08.r

#### ALIGNMENTS

RESULT 1  
A1079538/c  
LOCUS A1079538 536 bp mRNA linear EST 29-SEP-1998  
DEFINITION oz04e06.x1 Soares\_fetal\_liver\_spleen.INFLS.S1 Homo sapiens CDNA  
clone IMAGE:1674370 3', mRNA sequence.  
A1079538  
A1079538.1 GI:3415789

ACCESSION A1079538  
VERSION A1079538  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 536)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
This clone is available royalty-free through INFLS; contact the  
IMAGE Consortium ([info@image.jnl.gov](mailto:info@image.jnl.gov)) for further information.  
Insert Length: 1447 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 445.  
Location/Qualifiers

#### FEATURES

source

1..536  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1674370"  
/clone\_lib="Soares\_fetal\_liver\_spleen.INFLS.S1"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)  
with a modified polylinker; Site1: Pac I; Site2: Eco RI;  
This is a subtracted version of the original Soares fetal  
liver spleen INFLS library. 1st strand cDNA was primed  
with a Pac I - oligo(dT) primer [5'  
AACGGAAGATTAATTAACATCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

BASE COUNT 174 a 126 c 114 g 122 t

ORIGIN

Query Match 100.0%; Score 249; DB 9; Length 536;  
Best Local Similarity 100.0%; Pred. No. 1,1e-51;  
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 cagcggtgtgtgcaggaagctcactctcgctcagatattagatgtgtgtgtgtcgcg 60  
|||  
Db 497 CAGCGGTGTGTGCGCAGGAAGCTCCTCGCTCAGTATTAGATGTGTGTGTGTGCG 438  
|||  
Y 61 gggatctgtgtgtccatctcctcattgttctgtgaacatctgtatgtaaccatg 120  
|||  
Db 437 GGGATCTGCGTGTGCGCAGGAAGCTCCTCGCTCAGTATTAGATGTGTGTGTGTGCG 378  
|||  
Y 121 gctgggtgtcctaaagctcgtcgtgaatcccgatgtggaagctgtgaagctcag 180  
|||  
Db 377 GCTGGGTGTCTAAAGTGTCTGTGATCCCGATGTGGAAGCTGTGGAAGCTCAG 318  
|||  
Y 181 catccatgtattactttaaaacagaaaaagacatgatatgtatgtctattttt 240  
|||  
Db 317 CATACCATGTATTACTTTAAACAGAAAAAGACATGTATGTATGTATTTT 258  
|||  
Y 241 ttattatgg 249  
|||  
Db 257 TTTTATTGG 249

RESULT 2  
AI497808 536 bp mRNA linear EST 14-APR-1999  
LOCUS tm89f02.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2165307 3',  
DEFINITION mRNA sequence.  
ACCESSION AI497808  
VERSION AI497808.1 GI:4389790  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 536)  
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BRTP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.llnl.gov/bdrrp/image/image.html  
Insert Length: 779 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 468.

#### FEATURES

Location/Qualifiers  
1..536  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2165307"  
/clone\_1lb="NCI CGAP Brn25"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"

/note="Organ: brain; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'] TGTACCATCTGAGTGGAGCGGCGCCATAGGTTTTTTTTTTTTTTTTTTT T 3'; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 174 a 126 c 114 g 122 t

ORIGIN

Query Match 100.0%; Score 249; DB 9; Length 536;  
Best Local Similarity 100.0%; Pred. No. 1,1e-51;  
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 cagcggtgtgtgcaggaagctcactctcgctcagatattagatgtgtgtgtgtcgcg 60  
|||  
Db 497 CAGCGGTGTGTGCGCAGGAAGCTCCTCGCTCAGTATTAGATGTGTGTGTGTGCG 438  
|||  
Y 61 gggatctgtgtgtccatctcctcattgttctgtgaacatctgtatgtaaccatg 120  
|||  
Db 437 GGGATCTGCGTGTGCGCAGGAAGCTCCTCGCTCAGTATTAGATGTGTGTGTGTGCG 378  
|||  
Y 121 gctgggtgtcctaaagctcgtcgtgaatcccgatgtggaagctgtgaagctcag 180  
|||  
Db 377 GCTGGGTGTCTAAAGTGTCTGTGATCCCGATGTGGAAGCTGTGGAAGCTCAG 318  
|||  
Y 181 catccatgtattactttaaaacagaaaaagacatgatatgtatgtctattttt 240  
|||  
Db 317 CATACCATGTATTACTTTAAACAGAAAAAGACATGTATGTATGTATTTT 258  
|||  
Y 241 ttattatgg 249  
|||  
Db 257 TTTTATTGG 249

RESULT 3  
AI038061 561 bp mRNA linear EST 28-AUG-1998  
LOCUS ox21b05.x1 Soares\_fetal\_liver\_spleen\_JNFs\_S1 Homo sapiens cDNA  
DEFINITION clone IMAGE:1656945 3', mRNA sequence.  
ACCESSION AI038061  
VERSION AI038061.1 GI:3277255  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 561)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1466 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 465.

#### FEATURES

Location/Qualifiers  
1..561  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1656945"  
/clone\_1lb="Soares\_fetal\_liver\_spleen\_JNFs\_S1"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;





High quality sequence start: 20  
High quality sequence stop: 617.

FEATURES  
source

Location/Qualifiers  
1. .618  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0585"  
/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pOC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 178 a 146 c 124 g 170 t  
ORIGIN

Query Match 97.4%; Score 242.6; DB 9; Length 618;  
Best Local Similarity 98.4%; Pred. No. 4.2e-50;

Matches 245; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cagcgggtggtgcaaggaagctcactctgcgcagatattagatgtgtgtgtgtctcg 60  
|||  
Db 303 CAGCGGTGTCGCAAGAGCTCAGCTCGCGCTAGATAGAGTGTGTGTGTGTCTCG 244  
QY 61 gggatctgggtggtccatctctctcattgttgaacatcctgtatgaacatcg 120  
|||  
Db 243 GGGATCTGGGTGGTCCATCTCTCTCATTTCTTGAACATCTGTATTTGAACCATG 184  
QY 121 gctgggtgtcctaaagtgcctgtgaatccgatgtgaaagctgaggtgaaagctcag 180  
|||  
Db 183 GCTGGGTGTCAAAAGTGTGCTGATCCGATGTGAAAGCTGAGGTGAAGCTCG 124  
QY 181 cataccatgtatttactttaaaacagaaaaaagacatgtatgtatgtctattttt 240  
|||  
Db 123 CATACCATGTATTACTTTAAATACAGAAAAAGACATGTATGTATGTATTTT 64  
QY 241 ttattatgg 249  
|||  
Db 63 TTTTACTCG 55

## RESULT 6

AM242451 540 bp mRNA linear EST 14-DEC-1999

LOCUS XM939404.x1 NCI.CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2692302 3',  
mRNA sequence.

ACCESSION AM242451  
VERSION AM242451.1 GI:6576205

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 540)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapsb-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

CNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ULNL at:

www.bio.lnlnl.gov/dbip/image/image.html  
www-bio.lnlnl.gov/dbip/image/image.html

Seg primer: -400p from Gibco  
High quality sequence stop: 463.

FEATURES  
source

Location/Qualifiers  
1. .540  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image:2692302"

/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clones 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 175 a 126 c 113 g 126 t  
ORIGIN

Query Match 96.8%; Score 241; DB 9; Length 540;  
Best Local Similarity 98.0%; Pred. No. 1.1e-49;

Matches 244; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 cagcgggtggtgcaaggaagctcactctgcgcagatattagatgtgtgtgtgtctcg 60  
|||  
Db 501 CAGCGGTGTCGCAAGAGCTCAGCTCGCGCTAGATAGAGTGTGTGTGTGTCTCG 442  
QY 61 gggatctgggtggtccatctctcattgttgaacatcctgtatgaacatcg 120  
|||  
Db 441 GGGATCTGGGTGGTCCATCTCTCTCATTTCTTGAACATCTGTATTTGAACCATG 382  
QY 121 gctgggtgtcctaaagtgcctgtgaatccgatgtgaaagctgaggtgaaagctcag 180  
|||  
Db 381 GTTGGGTGTCTTAAAGTCCCTGTGAATCCGATGTGAAAGCTGAGGTGAAGCTCAG 322  
QY 181 cataccatgtatttactttaaaacagaaaaaagacatgtatgtatgtctattttt 240  
|||  
Db 321 CATACCATGTATTACTTTAAAAACAGAAAAAGACATGTATGTATGTATTTT 262  
QY 241 ttattatgg 249  
|||  
Db 261 TTTTATTGG 253

## RESULT 7

AI333325 486 bp mRNA linear EST 13-FEB-1999

LOCUS q905b01.x1 Soares.NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1931593  
DEFINITION 3', mRNA sequence.

ACCESSION AI333325  
VERSION AI333325.1 GI:4069884

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 486)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapsb-r@mail.nih.gov

This clone is available royally-free through ULNL; contact the  
IMAGE Consortium (info@image.lnlnl.gov) for further information.  
Insert Length: 782 Std Error: 0.00

Seg primer: -400p from Gibco  
High quality sequence stop: 447.

Location/Qualifiers  
1. .486  
/organism="Homo sapiens"

```

/ldb Xref="TAXON:9606"
/clone_image="1931593"
/clone_lib="Soares-NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below): Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of T.M.A.G.E. clones 260232-265223,
340488-343479, and 484488-489479."

```

Query Match	95.2%	Score 237;	DB 9;	Length 486;
Best Local Similarity	100.0%	Pred. No. 1.1e+48;		
Matches 237; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	13	caggaagctcactctcgacgaattagaatgtgtgtgagtcctggagactcctggtg	72
Db	486	CAGGAAGCTCACTCTCGCGTGAATTAGAAATGTGTGTGTGGCTCGGGGAACTCCGGT	427
QY	73	gtccccalcctccctcatctgtctcgaacatccctgtatgttaaacatcagctcgtgggtcta	132
Db	426	GCTCCCACTCTCTTCATATGTGTGAAACATCTGTATGTAAACAATGCGCTGGGGTCTA	367
QY	133	aagtgctcgttaatcccgatgtggaataaaagctgtagagtgaaagctcagcatacacatgtat	192
Db	366	AAGTGCCGTAAATCCCGATGTGAAATAAAGCTGGAGGCTGAAGCTCAGCATACCATGTAT	307
QY	193	ttaactaaanaacgaaaanaagacatgtatgagtaagtcataatcttttttttttttgg	249
Db	306	TTACTTTAAACAAGAAAANAAGCATGTATGGAATATGTCATATTTTATTTTATATGG	250

	RESULT	8
AI952086/c	LOCUS	
DEFINITION	AI952086	486 bp mRNA linear EST 06-SEP-1999
ACCESSION	wk46g91.1.x1 NCI-CGAP_Lu28	Homo sapiens cDNA clone IMAGE:2546756 3'
VERSION	AI952086	
KEYWORDS	AI952086.1 GI:5744396	
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
COMMENT	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
CONTACT	Tumor Gene Index	
	Unpublished (1997)	
	Contact: Robert Strausberg, Ph.D.	

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Chris Mosekaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bdrrp/image/image.html](http://www-bio.llnl.gov/bdrrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 413.  
Location/Qualifiers  
1.486

```

-
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2346756"
/clone_1ib="NCI.CGAP.Lu28"
/tissue_type="two pooled squamous cell carcinomas"
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-Sport6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT
library constructed by Life Technologies."
BASE COUNT      164 a      112 c      100 g      110 t
ORIGIN

```

Query Match	95.2%;	Score 237;	DB 9;	Length 486;
Best Local Similarity	100.0%;	Pred. NO. 1.1e-48;		
Matches 237; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0

Oy	13	caggaagctcaactctcgacgaatcaagctagagctgctgtgctggagctcgcgtg	72
Db	486	CAGGAAGCTCACTCGCGTCAAGTATTAGAGCTGCTGTGCGGCTCGGGGACTCGCGTG	427
Oy	73	gctccacatcttccttcattgcttctgaacatcctgtaattgaacaacagctgagtgcta	132
Db	426	GCTCCACATCTTCTTCAATTGTCTGAAACATCCTGTATTGTAAACAGCTCGGGGTGCTA	367
Oy	133	aagtgctctgtaatacccgatgtgtaaaaaagctggaggtgaagctcagcatcacatgatat	192
Db	366	AAGTGCTCTGAATCCGAGTGTGAAAAAAGCTGGAGGCTGAACCTCAGCATACCATGTAT	307
Oy	193	ttaacttaaaaaagaaaaaagacatgctatgatactatctatctttttttttttttttttg	249
Db	306	TTAACTTTAAAACGAAAAAAGACATGTAATGATATGCTATTTTTTTTTTTTATGG	250

[illegible]

```

FEATURES
SOURCE

location/Qualifiers
1..545
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1713028"
/clone_1id="Scares-pregnant_uterus_NBDPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AAGCGAGGAATTCGCGCGCCCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

```

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"
```

```
1. .463
/organism="Homo sapiens"
```

[illegible]

	/clone="IMAGE:1913978"	
	/clone_lib="NCI_CGAP_Kid5"	
	/tissue_type="2 pooled tumors (clear cell type)"	
	/lab_host="DHIOB"	
	/note="Organ: kidney; Vector: pTf73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I... oligo(dT) primer [5' AATCGAGAAGATTCTGGCCGCACATATTTTTTTT TTTT TTTT TTTT double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified Pf7n3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."	
BASIC COUNT	157 a    105 c         92 g              108 t	
ORIGIN		
Query Match	83.5%; Score 208; DB 9; Length 462;	
Best Local Similarity	100.0%; Pred. No. 1.7e-41;	
Matches 208:	Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
OY	42 agtggtgctggagcgcgatcgatgctcccatcttcctcaattgttcgaaca 101	
Dd	462 AGTGGTGTGTGGGTCCTGGGGATTCGGGTGCCCATCTTCTCATTGTTCGACA 403	
OY	102 tcctgtatgttaaacaccatgcgtgggtgcltaaagtacctgtgaatacccgaatgtgaaaaa 161	
Dd	402 TCCTGTATGTAAACCATGCGCTGGGGTCTTAAGTGGCTGTGAATCCGATGTGAAAAA 343	
OY	162 gcctggaggtagaaactcacatcacatgatatcttatctaataaaaagaagacatgt 221	
Dd	342 GCTGAGAGTAGAACGTCAGCAATCACCTATATTACTTTAAAACAGAAAAAGACATGT 283	
OY	222 atggatgatctatcttttttttttgtg 249	
Dd	282 ATGGATGATGCTATTTTTTTTTTTATTGG 255	
RESULT 13		
A1369277/c		
LOCUS		
DEFINITION	A1369277 446 bp mRNA linear EST 17-MAR-1997	
ACCESSION	Gy91bl1.x1 NCI_CGAP_Brn25 Homo sapiens CDNA clone IMAGE:2019357 3'	
VERSION	A1369277	
KEYWORDS	mRNA sequence.	
SOURCE	EEST.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 446)	
TITLE	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/hic/gap. National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP). Tumor Gene Index Unpublished (1998) Contact: Robert Strausberg, Ph.D. Email: cgabsr@mail.nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. CDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. CDNA library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LBLW at: <a href="http://www.bio.linal.gov/bdrp/image/image.html">www.bio.linal.gov/bdrp/image/image.html</a> Insert length: 723 Std Error: 0.00 Seq primer: -40Up from GlBCO. Location/Qualifiers 1..446 /organism="Homo sapiens" /db_xref="taxon:9606"	
JOURNAL COMMENT		
FEATURES		
source		

```

/clone="IMAGE:2019357"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGATTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

```

Query Match	79.5%	Score 198	DB 9	Length 446
Best Local Similarity	100.0%	Pred. No. 5	le-39	
Matches 198	Conservative 0	Mismatches 0	Indels 0	Gaps 0

  

Oy	52	tgggtctcggggagcctcgtggtcccatcttcctcattgltctgacatccgtatg	111
Db	446	tgggctctgggggattctcggtgctcccatcttcccttattgttctgaaacatcctgattg	387
Oy	112	taaacatcggctcgggggtgtctaagtcgctgtgaatcccgatgtggaataaagctggaggtc	171
Db	386	taaacacatggcctggggctgctaaagtcctgtaaacctcgtataatcccatgtggaaaaagctggaggtg	327
Oy	172	aaagctcagcatatcatgttatttactttaaacaagaaaaaagacatgtcatgatatgt	231
Db	326	aaagctcagcatatccatgatttacttttaaaaaagaaaaaagacatgtatgatatgt	267
Oy	232	ctattttttttttttt	249
Db	266	ctatttttttttttttgg	249

```

RESULT 14
A1192867/c
LOCUS
DEFINITION
A1192867 436 bp mRNA linear EST 29-OCT-1996
gene8b05.x1 Soares_fetal_lung_NBHL19w Homo sapiens cDNA clone
IMAGE:1744113 3', mRNA sequence.
ACCESSION
A1192867
VERSION
A1192867.1 GI:3744076
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 436)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LML; contact the
IMAGE Consortium (lmfimage.llnl.gov) for further information.
Insert Length: 798 Std Error: 0.00
Seq primer: -40UP from Glenco.
FEATURES
source
1..436
Location/Qualifiers

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1744113"
/clone_lib="Soares fetal_lung_NbHL19"
/dev_stage="19 weeks"
/lab_host="DH10b (ampicillin resistant)"
/name="Organ:lung;Vector: pT73D (Pharmacia) with a
modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
5'-TGTTACCAATCTAATGGGAGCGGCCGCAATTTTGTGTTTTTTT-3'

```

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fátima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NDH19W."

	Query Match	75.1%	Score 187	DB 9	Length 436
	Best Local Similarity	100.0%	Pred. No. 2,7e-36		
	Matches 187	Conservative 0	Mismatches 0	Indels 0	Gaps 0
OY	63	gactcgtgtgctccacatcttccttcattgtlctgaaacatcctgattgttaaacatgac	122		
Db	436	GATCTCGGTGGTGGTCCCATCTCTTCATGTGTTCTCAACATCTCTGATTTTAACCATGGC	377		
OY	123	tggagttgtaagtcgtctgtaatccgtagtggaaagctggaggtttaaagctcaga	182		
Db	376	TGGGTTGCTAAAGTCCCTGTGAATCCCATGTGGAAAAAGCTGGAGGTAAAGCTACGA	317		
OY	183	taccatgattactttaaaaaacagaaaaaagacatgtafgatagtcatactttttt	242		
Db	316	TACCATGATTTTACTTTAAAAACAGAAAAAAGACATGATGATATGCTATTTTTTT	257		
OY	243	ttatgtg 249			
Db	256	TTATTGG 250			

RESULT	15
AA293275/c	
LOCUS	
DEFINITION	427 bp mRNA linear EST 08-AUG-1997 AA293275 Z138e06.s1 Soares ovary tumor NBHOT Homo sapiens cDNA clone IMAGE:172642.3 similar to contans element L1 repetitive element , mRNA sequence.
ACCESSION	AA293275
VERSION	AA293275
KEYWORDS	GI:1941425
SOURCE	EST. human.
ORGANISM	Homo sapiens
REFERENCE	Eukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 427)
AUTHORS	Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissee,S., Dietrich,N., Dubugue,T., Favello,A., Gish,W., Hawkins, M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Nardis,E., Moore, B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaski,E., Underwood,K., Wohlmann,P., Watson,R., Wilson,R. and Marra,M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996) 97044478
TITLE	
JOURNAL	
MEDLINE	
COMMENT	Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGS Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1397 Std Error: 0.00  
Seq primer: -41m3 fwd. ET from Amersham  
High quality sequence stop: 395.  
Location/Qualifiers  
FEATURES  
source  
1..427  
/organism="Homo sapiens"  
/db\_xref="GDB:593626"  
/db\_xref="taxon:9606"  
/clone="IMAGE:724642"



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 04:48:14 ; Search time 215.36 Seconds  
(without alignments)  
1985.103 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 249  
Sequence: 1 cagcggtgtgtgcaggaagc.....gtctatctttttttatcttg 249

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	249	100.0	867	22	AAF91875	Human secreted pro
2	249	100.0	5510	22	AAH73337	Human cervical can
3	36.2	14.5	7784	24	ABL34424	Human immune syste
4	36	14.5	2529	18	AAH7468	Hamster Ubiqutin/B
5	35.4	14.2	151826	21	AAH22291	BAC containing rep
6	35.2	14.1	619	21	AAZ80495	Human colon cancer
7	34.8	14.0	1734	22	AAH17416	Human CDNA sequenc
8	34.8	14.0	2364	21	AAH70246	Plasmodium falcipa
9	34.4	13.8	13416	24	ABL34231	Human immune syste

10	34.4	13.8	15416	24	AAH61453	Human gene regulat
11	34.2	13.7	7784	24	ABL34425	Human immune syste
12	34	13.7	1900	20	AAZ25197	Nucleotide sequenc
13	34	13.7	6298	22	AAH45358	Chemically pretrea
14	33.8	13.6	540	22	AAH10129	Human CDNA clone (
15	33.8	13.6	5536	24	ABL32179	Human immune syste
16	33.8	13.6	7784	24	ABL34424	Human immune syste
17	33.6	13.5	13104	19	AAV52167	Streptococcus pneu
18	33.6	13.5	11000	22	AAH48800	Nucleotide sequenc
19	33.4	13.4	1267	22	AAH21012	Bovine-derived DNA
20	33.4	13.4	6306	22	AAH85746	Human immune/haema
21	33.4	13.4	8395	20	AAH31354	Enterococcus faeca
22	33	13.3	881	17	AAH40048	Human tumour necro
23	33	13.3	881	21	AAH28149	Human TR2 receptor
24	33	13.3	1704	19	AAV34509	Human TNF receptor
25	33	13.3	1704	20	AAH83763	Human tumour necro
26	32.8	13.2	265	17	AAH28133	Senescence-related
27	32.8	13.2	8032	24	AAH61408	Human gene regulat
28	32.6	13.1	5914	24	ABL34169	Human immune syste
29	32.6	13.1	6398	23	ABL13526	Drosophila melanog
30	32.4	13.0	5062	22	AAH46696	Tumour suppressor
31	32.2	12.9	500	21	AAH3256	Cat flea head and
32	32.2	12.9	1804	23	AAH83281	DNA encoding novel
33	32.2	12.9	3069	22	AAH54988	S. epidermidis gen
34	32.2	12.9	3206	22	AAH55018	S. epidermidis gen
35	32.2	12.9	4249	22	AAH54769	S. epidermidis gen
36	32.2	12.9	5536	24	ABL32178	Human immune syste
37	32.2	12.9	6351	24	ABL33745	Human immune syste
38	32.2	12.9	6351	24	ABL34587	Human immune syste
39	32	12.9	5536	24	ABL32179	Human metastasis a
40	32	12.9	17341	21	AAH14872	Human immune syste
41	32	12.9	32150	22	AAH32454	Genomic DNA sequen
42	31.8	12.8	407	22	AAH20983	Human genomic DNA
43	31.8	12.8	407	22	AAH69467	Human nervous syst
44	31.8	12.8	1383	24	AAH95277	Human digestive sy
45	31.8	12.8	5314	24	ABL32160	Long terminal repe
						Human immune syste

#### ALIGNMENTS

RESULT 1	
AAF91875	AAF91875 standard; CDNA; 867 BP.
XX	XX
AC	AAF91875;
XX	XX
DT	22-MAY-2001 (first entry)
XX	XX
DE	Human secreted protein-encoding gene 18 CDNA clone HEPFK57, SEQ ID NO:28.
XX	XX
KW	Human; secreted protein; proliferative disorder; cancer; tumour;
KW	focal abnormality; developmental abnormality; haematopoietic disorder;
KW	immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;
KW	inflammation; allergy; neurological disorder; Alzheimer's disease;
KW	Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW	cardiovascular disorder; angiotensin disorder; kidney disorder;
KW	gastrointestinal disorder; pregnancy-related disorder;
KW	endocrine disorder; infection; wound healing; vulnerability;
KW	cell culture; chemotaxis; food additive;
KW	binding partner identification; ss.
OS	XX
OS	Homo sapiens.
XX	XX
PN	WO200118022-A1.
XX	XX
PD	15-MAR-2001.
XX	XX
PF	31-AUG-2000; 2000OMO-US24008.
XX	XX
PR	03-SEP-1999; 99US-0152315.
PR	03-SEP-1999; 99US-0152317.



XX (HUMA-) HUMAN GENOME SCI INC.  
PA

XX NI J. Baker KP, Birse CE, Fiscella M, Komatsu G, Rosen CA;  
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;  
PI Moore PA, Shi Y, Wei Y, Florence KA;  
XX

DR WPI: 2001-203081/20.  
DR P-PSDB; AAB87359.  
XX

PT Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers -  
XX

PS Claim 1; Page 498; 607pp; English.  
XX

CC AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted  
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.  
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes  
CC and their corresponding secreted proteins are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 52 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein-encoding cDNA of the invention.  
XX

SQ Sequence 867 BP; 233 A; 180 C; 197 G; 256 T; 1 other;

Query Match 100.0%; Score 249; DB 22; Length 867;  
Best Local Similarity 100.0%; Pred. No. 2.9e-65;  
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcgggtggtgaggaagtcactcgcgtcagatagagtggtgtggtctcg 60  
DB 348 cagcgggtggtgaggaagtcactcgcgtcagatagagtggtgtggtctcg 407  
QY 61 gggatctcgtggtcccatcttcctcattgtctgaaacatccgtatgtgaaacatg 120  
DB 408 gggatctcgtggtcccatcttcctcattgtctgaaacatccgtatgtgaaacatg 467  
QY 121 gctgggggtgctaaagtgccctgtgaatcccgatgtggaagcgtggaagcctcag 180  
DB 468 gctgggggtgctaaagtgccctgtgaatcccgatgtggaagcgtggaagcctcag 527  
QY 181 cataccatgtattactttaaacaacagaaagacatgtatgagatgtctattttt 240  
DB 528 cataccatgtattactttaaacaacagaaagacatgtatgagatgtctattttt 587  
QY 241 ttattatgg 249  
DB 588 ttattatgg 596

RESULT 2

AAH73337  
ID AAH73337 standard; cDNA; 5510 BP.  
XX

AAH73337;  
XX

19-SEP-2001 (first entry)  
XX

Human cervical cancer marker nucleic acid 4611.  
XX

Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX

Homo sapiens.  
XX

WO200142467-A2.  
XX

14-JUN-2001.  
XX

08-DEC-2000; 2000WO-US33312.  
XX

08-DEC-1999; 99US-0169681.  
XX

21-DEC-1999; 99US-0171350.  
XX

14-MAR-2000; 2000US-0189315.  
XX

12-MAY-2000; 2000US-0203791.  
XX

09-JUN-2000; 2000US-0210600.  
XX

21-JUL-2000; 2000US-0220114.  
XX

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Deeds J, Berger A, Zhao X;  
XX

WPI: 2001-375006/39.  
XX

New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
XX

Claim 1; Page 1029-1031; 1051pp; English.  
XX

XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
XX cervical cancer with cytostatic activity. The nucleic acids and encoded  
XX polypeptides are useful to assess if a patient is afflicted with  
XX cervical cancer or has a pre-malignant condition; to monitor the  
XX progression of cervical cancer or a premalignant condition in a patient;  
XX and to select and/or assess the efficacy of a compound or therapy for  
XX inhibiting cervical cancer in a patient. The nucleic acids may also be  
XX useful for gene therapy.

SQ Sequence 5510 BP; 1526 A; 1242 C; 1418 G; 1305 T; 19 other;

Query Match 100.0%; Score 249; DB 22; Length 5510;  
Best Local Similarity 100.0%; Pred. No. 5.7e-65;  
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcgggtggtgaggaagtcactcgcgtcagatagagtggtgtggtctcg 60  
DB 4914 cagcgggtggtgaggaagtcactcgcgtcagatagagtggtgtggtctcg 4973  
QY 61 gggatctcgtggtcccatcttcctcattgtctgaaacatccgtatgtgaaacatg 120  
DB 4974 gggatctcgtggtcccatcttcctcattgtctgaaacatccgtatgtgaaacatg 5033  
QY 121 gctgggggtgctaaagtgccctgtgaatcccgatgtggaagcgtggaagcctcag 180  
DB 5034 gctgggggtgctaaagtgccctgtgaatcccgatgtggaagcgtggaagcctcag 5093  
QY 181 cataccatgtattactttaaacaacagaaagacatgtatgagatgtctattttt 240  
DB 5094 cataccatgtattactttaaacaacagaaagacatgtatgagatgtctattttt 5153  
QY 241 ttattatgg 249  
DB 5153 ttattatgg 519







CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
XX specifically mentioned within the specification.  
SO Sequence 2364 BP; 1222 A; 247 C; 213 G; 682 T; 0 other;

Query Match 14.0%; Score 34.8; DB 21; Length 2364;  
Best Local Similarity 65.4%; Pred. No. 1.7;  
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 170 tgaagtcagcatcacatgtattacttaaaacagaaaaaagacatgatgatgat 229  
DB 1854 TGAATATATCTGCATGACATGACTTTAATTATATCATCAACAATATTTTGTATAT 1795

OY 230 gtctatttttttttatt 247  
DB 1794 ATGTGAATATATTTTATT 1777

RESULT 9  
ABL34231  
ID ABL34231 standard; DNA; 15416 BP.

XX ABL34231;  
XX  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Human immune system associated gene SEQ ID NO: 2204.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
XX antiarteriosclerotic; antianaemic; cytosolic; noctropic;  
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
XX antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
XX gene; ds.

XX Homo sapiens.  
XX  
XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
XX for diagnosis and treatment of diseases associated with abnormal  
XX cytosine methylation

XX Claim 1; SEQ ID NO 2204; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
XX can be used in the diagnosis and treatment of immune system disorders,  
XX including eye diseases such as retinopathy, neovascular glaucoma and  
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
XX diseases. The present sequence is a gene of the invention.

XX Sequence 15416 BP; 4742 A; 239 C; 3068 G; 7367 T; 0 other;

Query Match 13.8%; Score 34.4; DB 24; Length 15416;  
Best Local Similarity 57.4%; Pred. No. 4.5;  
Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 140 tgtgaccgcatgtgtaaaagcgtgaagtcagcatcacatgtattacttt 199  
DB 1966 tttatttttgatattgtatgaattataataaagtagatgaatgaatgaatgaatt 2025

OY 200 aaaaacagaaaaaagacatgatgatgtctatttttttttttatt 247  
DB 2026 tataaagaataaagtagatgaatgaatgttttttttttttttttttttttttttttt 2073

RESULT 10  
AAS61453  
ID AAS61453 standard; DNA; 15416 BP.

XX AAS61453;

XX 29-JAN-2002 (first entry)

XX Human gene regulation-associated gene oligonucleotide #408.

XX Human; gene regulation-associated gene; severe combined immunodeficiency;  
XX cardiac damage; inflammatory response; haemophilia; Werner syndrome;  
XX asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
XX renal disease; Preeclampsia; cardiac allograft vascular disease;  
XX colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
XX immunostimulant; cardiact; antineoplastic; coagulant; antiasthmatic;  
XX nephrotropic; gynecological; anti-tumour; immunosuppressive; cytosolic.

XX Homo sapiens.  
XX  
XX WO200177375-A2.

XX 18-OCT-2001.

XX 06-APR-2001; 2001WO-EP03968.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-017470/02.

XX New nucleic acid sequences from chemically modified genes associated  
XX with gene regulation, useful for analysing cytosine methylations for  
XX diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
XX disease

XX Disclosure; SEQ ID NO 418; 26pp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least  
XX 18 bases of a chemically pretreated gene associated with gene regulation  
XX selected from 43 known genes (or complementary sequences). The  
XX chemical pretreatment converts cytosine bases unmethylated at the  
XX 5-position to uracil or another base with hybridisation behaviour  
XX dissimilar to cytosine, to enable analysis of cytosine methylations.  
XX The DNA sequences, oligomers (or sets/arrays) and method are

XX useful in the diagnosis of diseases (or predisposition to diseases)  
XX associated with gene regulation and in therapy of such diseases, by  
XX enabling analysis of the cytosine methylation patterns of such genes,  
XX kits are provided. They are especially useful in diagnosis  
XX and therapy of e.g. severe combined immunodeficiency disease, cardiac  
XX disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
XX asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
XX Preeclampsia, graft versus-host disease. The present sequence is a





CC represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 540 BP; 160 A; 123 C; 137 G; 116 T; 4 other; SQ

Query Match	13.6%	Score 33.8	DB 22	Length 540	-
Best Local Similarity	53.0%	Pred. No. 2			
Matches 71	Conservative	0	Mismatches 63	Indels 0	Gaps 0

**DY**    115 accatgctgcgggttgctaagaagcctgtgaatcccgatttggaaaaaacgttaggtgaaa 174  
       ||| | | | | | | | | | |  
**Db**      442 ACCTCGATGGAGCTTCA TGAGA CTGTGAATAACTNCAGTAACCCAGATTGCCAAGAAG 383

QY 175 gctcagcataccatgctatttactttaaaaaacagaatacgtatgatgtccta 234  
| | | | | | | | | | | | | | | |  
Db 382 CCAGAGTCGAGTTATAGTACTTCAGACCTCGATTTAAGAATTCCTTTTCTCCT 323

Qy	235	ttttttttcatg	248
Db	322	TTTTTTTTTTTTTG	309

RESULT 15  
ABL32179

XX  
AC ABL32179;  
XX

Human immune system associated gene SEQ ID NO: 152.

[illegible]

acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene; ds.

aa  
os  
xx  
PN

Homo sapiens.

W0200200928-A2.

AA 03-JAN-2002.  
PD  
XX  
PF 02-JUL-2001: 2001WO-EP07537.

AA 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX

FA (EP10-) EP10ENOMILS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K.  
XX

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT  
UR WPI; 2002-130909/1/.

XX	Sequence
5536 BP; 1644 A; 86 C; 1202 G; 2604 T; 0 other	

Query Match	13.6%	Score 33.8	DB 24	Length 5536
Best Local Similarity	77.4%	Pred. No. 4.7		
Matches 41	Conservative 0	Mismatches 12	Indels 0	Gaps 0

QY 192 ctactcttaaaacagaaaaaaaagacatgatgatatgtctattttttt 244  
||| ||||| | | ||||| | | ||| | ||||| |||||  
Db 670 tttttttaagaagaaaaaaaataataataataataataataattttttt 722

Search completed: September 20, 2002, 05:59:35



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 04:01:20 : Search time 2113.79 seconds  
(without alignments)  
2465.102 Million cell updates/sec

Title: US-09-757-781-21  
Perfect score: 249  
Sequence: 1 cagcggtggtgagcaggaagc.....gtctattttttttttattg 249

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: gb\_ba:\*
  - 2: gb\_hgt:\*
  - 3: gb\_in:\*
  - 4: gb\_om:\*
  - 5: gb\_ov:\*
  - 6: gb\_pat:\*
  - 7: gb\_ph:\*
  - 8: gb\_pl:\*
  - 9: gb\_pr:\*
  - 10: gb\_ro:\*
  - 11: gb\_sy:\*
  - 12: gb\_sy:\*
  - 13: gb\_un:\*
  - 14: gb\_vl:\*
  - 15: em\_ba:\*
  - 16: em\_fun:\*
  - 17: em\_hum:\*
  - 18: em\_in:\*
  - 19: em\_mu:\*
  - 20: em\_om:\*
  - 21: em\_or:\*
  - 22: em\_ov:\*
  - 23: em\_pat:\*
  - 24: em\_ph:\*
  - 25: em\_pl:\*
  - 26: em\_ro:\*
  - 27: em\_sy:\*
  - 28: em\_un:\*
  - 29: em\_vl:\*
  - 30: em\_hgt\_hum:\*
  - 31: em\_hgt\_inv:\*
  - 32: em\_hgt\_other:\*
  - 33: em\_hgtgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
-----					

1	249	100.0	2696	9	AK000761	AK000761 Homo sapi
2	249	100.0	5958	9	AF196185	AF196185 Homo sapi
3	249	100.0	22279	2	U82210	U82210 Homo sapien
4	249	100.0	106198	9	AL160409	AL160409 Human DNA
5	249	100.0	138688	2	AC025823	AC025823 Homo sapi
6	249	100.0	196374	2	AL591464	AL591464 Homo sapi
7	176	70.7	427	6	AX332939	AX332939 Sequence
8	70	28.1	5500	10	AB005549	AB005549 Rattus no
9	50.6	20.3	168213	2	AC103175	AC103175 Rattus no
10	40.8	16.4	168280	2	AC103460	AC103460 Rattus no
11	39	15.7	2305	3	S96842	S96842 SP96-spore-
12	38.8	15.6	176741	9	AC016732	AC016732 Homo sapi
13	38.6	15.5	203194	2	AC023228	AC023228 Homo sapi
14	38.4	15.4	119579	2	AC008425	AC008425 Homo sapi
15	38.2	15.3	92795	2	AL357035	AL357035 Homo sapi
16	38.2	15.3	189655	2	AC103542	AC103542 Rattus no
17	38	15.3	197239	2	AC094815	AC094815 Rattus no
18	37.6	15.1	162260	2	AC021236	AC021236 Homo sapi
19	37.4	15.0	55774	2	AC091303	AC091303 Homo sapi
20	37.2	14.9	76113	2	AC023453	AC023453 Homo sapi
21	37	14.9	125301	9	HS388N15	Z99571 Human DNA S
22	36.8	14.8	12029	3	AE001412	AE001412 Plasmodiu
23	36.8	14.8	185182	9	AC093423	AC093423 Homo sapi
24	36.6	14.7	39608	9	AC009004	AC009004 Homo sapi
25	36.6	14.7	74470	9	AC079404	AC079404 Homo sapi
26	36.6	14.7	79085	9	AC006221	AC006221 Homo sapi
27	36.6	14.7	164309	30	AC067834	AC067834 Homo sapi
28	36.6	14.7	166500	9	AC007785	AC007785 Homo sapi
29	36.6	14.7	169572	2	AC108219	AC108219 Homo sapi
30	36.6	14.7	191830	2	AC025660	AC025660 Homo sapi
31	36.6	14.7	193351	9	AC025656	AC025656 Homo sapi
32	36.6	14.7	205908	8	AL669903	AL669903 Mus muscu
33	36.4	14.6	34722	2	SPBC16C6	AL021767 S. pombe c
34	36.4	14.6	36629	2	AC107279	AC107279 Rattus no
35	36.4	14.6	186135	2	AL669927	AL669927 Mus muscu
36	36.2	14.5	7784	6	AX347326	AX347326 Sequence
37	36.2	14.5	110000	2	AP002753	Continuation (3 of
38	36.2	14.5	110000	10	AE008684	Continuation (2 of
39	36.2	14.5	131972	9	AC008794	AC008794 Homo sapi
40	36.2	14.5	132090	10	AC005402	AC005402 Mus muscu
41	36.2	14.5	151558	9	AC021766	AC021766 Homo sapi
42	36.2	14.5	155633	2	AC078795	AC078795 Homo sapi
43	36.2	14.5	156060	2	AC004153	AC004153 Plasmodiu
44	36.2	14.5	169546	2	AC004157	AC004157 Plasmodiu
45	36.2	14.5	169794	2	AC004688	AC004688 Plasmodiu

ALIGNMENTS

RESULT 1	AK000761	2696 bp	mRNA	Linear	PRI 22-FEB-2000
LOCUS	AK000761				
DEFINITION	Homo sapiens cDNA FLJ20754 fis, clone HEP02246.				
ACCESSION	AK000761				
VERSION	AK000761.1 GI:7021050				
KEYWORDS	Oligo capping; fis (full insert sequence).				
SOURCE	Homo sapiens hepatoma cell_line:Hep2 cDNA to mRNA, clone_lib:HEP clone:HEP02246.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T., Suzuki,Y., Odayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T., and Sugano,S.				
TITLE	NEDO human cDNA sequencing project				
JOURNAL REFERENCE	Unpublished (2000)				
AUTHORS	2 (bases 1 to 2696)				
TITLE	Sugano,S., Suzuki,Y., Ota,T., Odayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.				
	Direct Submission				



Db 5582 GCTGGGGTCTAAAGTSCCTGATGATCCGATGTCGAAAGAGCTGAGGTAAGCTCAG 5641  
 QY 181 catcacatgtattactttaaaacagaaacacatgtatgatatgtctattttt 240  
 Db 5642 CATACCATGTATTACTTTAAACAGAAAAAAGACATGATGATATCTATTATTTT 5701  
 QY 241 ttatttttg 249  
 Db 5702 TTTTATTGG 5710

RESULT 3  
 LOCUS U82210 22279 bp DNA linear HTG 30-AUG-2001  
 DEFINITION Homo sapiens chromosome 10 clone CRI-JC2075 map 10p11.2, \*\*\*  
 ACCESSION U82210  
 VERSION U82210.1 GI:3983510  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_CANCELLED.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 22279)  
 AUTHORS Smith,D.R.  
 TITLE Sequencing of Human Chromosome 10  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 22279)  
 AUTHORS Du,L. and Smith,D.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-DEC-1996) Bioinformatics Division, Genome  
 Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA  
 COMMENT Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing  
 Technology. Data may contain low quality sequence and BAC/Cosmid  
 vector sequences.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 1 22279: contig of 22279 bp in length.  
 Location/Qualifiers  
 1. 22279  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /map="10p11.2"  
 /clone="CRI-JC2075"

BASE COUNT 5671 a 5238 c 5397 g 5959 t 14 others  
 ORIGIN

Query Match 100.0%; Score 249; DB 2; Length 22279;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-60;  
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcggatggtgcaggaagtcacatctcgcgtcagatattagaagtgtgtgtgtctcg 60  
 Db 10975 CAGCGGTGTTGCGAGACACTCCTCGCGTCAGTATTAGACTGTGTGTGGTCTCG 11034  
 QY 61 gggatctcgtgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctc 120  
 Db 11035 GGGATCTCGTGGCTCCATCTTCCTTCATTTGTCGAAACATCCTGATATTGTAACCATG 11094  
 QY 121 gctgggggtcctaagtgtcgtatcccgatgttgaaaaagcttgaggtgaaagctcag 180  
 Db 11095 GCTGGGGTCTAAAGTSCCTGATGATCCGATGTCGAAAGAGCTGAGGTAAGCTCAG 1154

QY 181 catcacatgtattactttaaaacagaaacacatgtatgatatgtctattttt 240  
 Db 11155 CATACCATGTATTACTTTAAACAGAAAAAAGACATGATGATATCTATTATTTT 11214  
 QY 241 ttatttttg 249  
 Db 11215 TTTTATTGG 11223

RESULT 4  
 LOCUS AL160409/c 106198 bp DNA linear PRI 18-JUL-2001  
 DEFINITION Human DNA sequence from clone RP11-406D17 on chromosome 10,  
 complete sequence.  
 ACCESSION AL160409  
 VERSION AL160409.12 GI:14970794  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 106198)  
 AUTHORS Sycamore,N.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 COMMENT requests: clonerequest@sanger.ac.uk  
 On Jul 19, 2001 this sequence version replaced gi:14625535.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em., EMBL; Sw.,  
 SwissProt; Tr., TrEMBL; Wp., WormPep; Information on the WormPep  
 database can be found at  
 http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence  
 was generated from part of bacterial clone compis of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-406D17 is from the library RP11-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACE3.6  
 IMPORTANT: This sequence is not the entire insert of clone  
 RP11-406D17 it may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true left end of clone RP13-348N17 is at 104199 in this  
 sequence. The true right end of clone RP11-490024 is at 2000 in  
 this sequence.  
 Location/Qualifiers  
 1. 106198  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-406D17"  
 /clone\_1fb="RP11-11.2"  
 27. 160  
 /note="MIR repeat: matches 3. 139 of consensus"  
 repeat\_region 989. 1039  
 /note="MIR repeat: matches 93. 144 of consensus"  
 repeat\_region 1232. 1275  
 /note="22 copies 2 mer ac 77% conserved"

repeat\_region 1344. .1651  
/note="Alusx repeat: matches 1. .302 of consensus"  
repeat\_region 4229. .4606  
/note="MIR1J repeat: matches 1. .403 of consensus"  
repeat\_region 4648. .4815  
/note="FRAM repeat: matches 4. .171 of consensus"  
repeat\_region 5215. .5516  
/note="Alusx repeat: matches 1. .300 of consensus"  
repeat\_region 6161. .6216  
/note="MIR repeat: matches 47. .104 of consensus"  
repeat\_region 6217. .6527  
/note="Alusx repeat: matches 1. .295 of consensus"  
repeat\_region 6538. .6637  
/note="MIR repeat: matches 104. .205 of consensus"  
repeat\_region 6646. .6950  
/note="Alub repeat: matches 1. .304 of consensus"  
repeat\_region 8061. .8354  
/note="Alub repeat: matches 1. .289 of consensus"  
repeat\_region 8443. .8602  
/note="AluJo/FRAM repeat: matches 152. .308 of consensus"  
repeat\_region 10005. .10072  
/note="L2 repeat: matches 2673. .2741 of consensus"  
repeat\_region 10542. .10849  
/note="MER57B repeat: matches 43. .386 of consensus"  
repeat\_region 10851. .10975  
/note="FRAM\_C repeat: matches 15. .133 of consensus"  
repeat\_region 11451. .11758  
/note="AluJo repeat: matches 1. .312 of consensus"  
repeat\_region 12992. .14459  
/note="pTR5 repeat: matches 504. .1874 of consensus"  
repeat\_region 14460. .14550  
/note="MER61E repeat: matches 377. .466 of consensus"  
repeat\_region 14551. .14771  
/note="pTR5 repeat: matches 1874. .8522 of consensus"  
repeat\_region 14772. .14923  
/note="Alusx repeat: matches 3. .154 of consensus"  
repeat\_region 14930. .15218  
/note="Alub repeat: matches 1. .293 of consensus"  
repeat\_region 15223. .15519  
/note="Alusq repeat: matches 1. .299 of consensus"  
repeat\_region 15661. .15818  
/note="MER5A repeat: matches 34. .189 of consensus"  
repeat\_region 15868. .15919  
/note="26 copies 2 mer aa 73% conserved"  
repeat\_region 16133. .16443  
/note="Alusx repeat: matches 1. .312 of consensus"  
repeat\_region 16599. .16707  
/note="L1MC4 repeat: matches 7728. .7834 of consensus"  
repeat\_region 16712. .16787  
/note="38 copies 2 mer ta 76% conserved"  
repeat\_region 16825. .17092  
/note="AluJo repeat: matches 1. .292 of consensus"  
repeat\_region 17094. .17254  
/note="AluJo/FRAM repeat: matches 134. .295 of consensus"  
repeat\_region 17579. .17889  
/note="Alusg repeat: matches 1. .306 of consensus"  
repeat\_region 18807. .19086  
/note="MIR1D repeat: matches 3. .238 of consensus"  
repeat\_region 19087. .19391  
/note="Alusq repeat: matches 1. .310 of consensus"  
repeat\_region 19392. .19542  
/note="MIR1D repeat: matches 238. .501 of consensus"  
repeat\_region 19608. .19904  
/note="AluJo repeat: matches 1. .295 of consensus"  
repeat\_region 20333. .20642  
/note="Alusx repeat: matches 1. .310 of consensus"  
repeat\_region 20670. .20956  
/note="Alub repeat: matches 1. .286 of consensus"  
repeat\_region 21297. .21601  
/note="L2 repeat: matches 2414. .2696 of consensus"  
repeat\_region 21857. .22175  
/note="AluJo repeat: matches 1. .311 of consensus"  
repeat\_region 22502. .22739

/note="Alusx repeat: matches 1. .239 of consensus"  
repeat\_region 23135. .23915  
/note="L1PA10 repeat: matches 5369. .6158 of consensus"  
repeat\_region 27500. .27629  
/note="L1MA3 repeat: matches 6171. .6304 of consensus"  
repeat\_region 28661. .28984  
/note="Alusq repeat: matches 1. .312 of consensus"  
repeat\_region 30001. .30132  
/note="66 copies 2 mer aa 62% conserved"  
repeat\_region 30233. .30328  
/note="48 copies 2 mer ta 65% conserved"  
repeat\_region 30365. .30644  
/note="L1MC4 repeat: matches 6633. .6940 of consensus"  
repeat\_region 30645. .31062  
/note="MIR repeat: matches 1. .424 of consensus"  
repeat\_region 31063. .31122  
/note="L1MC4 repeat: matches 6940. .6995 of consensus"  
repeat\_region 31133. .31547  
/note="L1MC4 repeat: matches 7167. .7597 of consensus"  
repeat\_region 31648. .32167  
/note="L1PA6 repeat: matches 5612. .6131 of consensus"  
repeat\_region 32275. .32552  
/note="AluJo repeat: matches 5. .282 of consensus"  
repeat\_region 32592. .32886  
/note="L1MC4 repeat: matches 7791. .7972 of consensus"  
repeat\_region 33154. .33272  
/note="MIR repeat: matches 80. .185 of consensus"  
repeat\_region 33283. .33415  
/note="AluJo repeat: matches 14. .153 of consensus"  
repeat\_region 33416. .34701  
/note="pTR5 repeat: matches 172. .2212 of consensus"  
repeat\_region 34713. .34840  
/note="AluJo/FRAM repeat: matches 178. .302 of consensus"  
repeat\_region 35365. .35654  
/note="AluJo repeat: matches 1. .288 of consensus"  
repeat\_region 36067. .36155  
/note="L2 repeat: matches 2622. .2710 of consensus"  
repeat\_region 36582. .36804  
/note="MER30 repeat: matches 1. .230 of consensus"  
repeat\_region 37547. .37852  
/note="Alusx repeat: matches 1. .311 of consensus"  
repeat\_region 37894. .38050  
/note="MER5A repeat: matches 9. .188 of consensus"  
repeat\_region 38447. .38744  
/note="Alusx repeat: matches 5. .302 of consensus"  
repeat\_region 39063. .39193  
/note="FRAM\_C repeat: matches 1. .131 of consensus"  
repeat\_region 39510. .39673  
/note="MER5A repeat: matches 4. .189 of consensus"  
repeat\_region 40932. .41102  
/note="L1PA7 repeat: matches 1497. .1667 of consensus"  
repeat\_region 41103. .41726  
/note="L1PA7 repeat: matches 5515. .6138 of consensus"  
repeat\_region 43386. .43747  
/note="THE1B repeat: matches 1. .364 of consensus"  
repeat\_region 44459. .44771  
/note="Alusq repeat: matches 3. .306 of consensus"  
repeat\_region 44991. .45302  
/note="AluJo repeat: matches 1. .304 of consensus"  
repeat\_region 45536. .51640  
/note="L1PA7 repeat: matches 11. .6145 of consensus"  
repeat\_region 51663. .51964  
/note="Alub repeat: matches 1. .308 of consensus"  
repeat\_region 51966. .52183  
/note="109 copies 2 mer aa 69% conserved"

Query Match 100.0% Score 249: DB 9: Length 106198:  
Best Local Similarity 100.0%: Pred. No. 2e-60:  
Matches 249: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
Oy 1 cagcggtgtgtgcaagagtcacactcgcgtcatatagagtgtgtcgtgctgcg 60  
|||||  
Db 88101 CAGCGGTGTGCGAAGAGTCACACTCTCGCGTCATATTAGAGTGTGTGTGGGCTCG 88042

```

OY      61  ggagctcgtggtcccatcttcctcattgtctcgaacatcctgtattgttaaccatg 120
          |||||||
Db      88041  GGGATCTCGGTGGCTCCATCTTCCTTCATGTTCTGAACATCGTATGTAAACCATG 87982
OY      121  gctgggggtcctaagtgctgtaacccgagtcgtaaaagctgtaagtgaaagctcag 180
          |||||||
Db      87981  GCTGGGGTCTTAAGTCTGTGATCCGATGTGAAAAAGCTGGAGGAGGAAAGCTCAG 87922
OY      181  cataccatgtattactcttaaaacagaaaaagacatgatagtatgtctattttc 240
          |||||||
Db      87921  CATACCATGTATTACTTTAAAAACAGAAAAAAGACATGTATGATATCTATTITTTT 87862
OY      241  tttattgtg 249
          |||||||
Db      87861  TTTTATTGG 87853

RESULT  5
AC025823 LOCUS      138688 bp      DNA      linear      HTG 24-JAN-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-60H16, WORKING DRAFT
ACCESSION  AC025823
VERSION     AC025823.6  GI:13811888
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 138688)
Smith,D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 138688)
Smith,D.R.
Direct Submission
Submitted (16-MAR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Apr 27, 2001 this sequence version replaced gi:9887686.
-----
Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
-----
Project Information
Center project name: hg317
-----
Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 133029 bases at least Q40
Consensus quality: 135140 bases at least Q30
Consensus quality: 136233 bases at least Q20
Insert size: 138187; sum-of-ctrls
Quality coverage: 6.3x in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1095: contig of 1095 bp in length
*
* 1096 1195: gap of unknown length
*
* 1196 2203: contig of 1008 bp in length
*
* 2204 2303: gap of unknown length
*
* 2304 9462: contig of 7159 bp in length
*
* 9463 9562: gap of unknown length
*
* 9563 25807: contig of 16245 bp in length

```

```

FEATURES
Source
1. 138688
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-60H16"
/clone_id="RP11"
1. 1095
/note="assembly_name:Contig14"
1196..2203
/note="assembly_name:Contig15"
2304..9462
/note="assembly_name:Contig55"
/clone_end:"7"
9563..25807
/note="assembly_name:Contig56"
clone_end:"SP6"
25808..39871
/note="assembly_name:Contig57"
39872..68325
/note="assembly_name:Contig58"
68426..138688
/note="assembly_name:Contig59"
BASE COUNT 41112 a 30000 c 29755 g 37221 t 600 others
ORIGIN
Query Match 100.0%; Score 249; DB 2; Length 138688;
Best Local Similarity 100.0%; Pred. No. 2; ie-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1  cagcggtggtgagcagaagctcctcgcgcagatagagtggtggtcgcg 60
          |||||||
Db      134742  CAGCGGTGAGGAGCAAGCTCAGCTCAGATTAAGTGTGTGGGCTCG 134683
OY      61  ggagctcgtggtcccatcttcctcattgtctcgaacatcctgtattgttaaccatg 120
          |||||||
Db      134682  GGGATCTCGGTGGCTCCATCTTCCTTCATGTTCTGAACATCGTATTGTAACCATG 134623
OY      121  gctgggggtcctaagtgctgtaacccgagtcgtaaaagctgtaagtgaaagctcag 180
          |||||||
Db      134622  GCTGGGGTCTTAAGTCTGTGATCCGATGTGAAAAAGCTGGAGGAGGAAAGCTCAG 134563
OY      181  cataccatgtattactcttaaaacagaaaaagacatgatagtatgtctattttc 240
          |||||||
Db      134562  CATACCATGTATTACTTTAAAAACAGAAAAAAGACATGTATGATATCTATTITTTT 134503
OY      241  tttattgtg 249
          |||||||
Db      134502  TTTTATTGG 134494

RESULT  6
AL591464 LOCUS      196374 bp      DNA      linear      HTG 19-DEC-2001
DEFINITION Homo sapiens chromosome 10 clone RP13-302D10, *** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
ACCESSION  AL591464.2  GI:17973978
VERSION     AL591464
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
Burton,J.
Direct Submission

```





Hollins, B., Homai, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louie, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Melker, M., Miner, G., Miner, Z., Mitchell, T., Mohabath, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tameris, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 168213)  
Morley, K.C.  
Direct Submission  
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Dec 21, 2001 this sequence version replaced gi:17062822.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GJFC  
Center clone name: CH230-98H12  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329P1st call to findhaplolist  
Consensus quality: 136914 bases at least Q40  
Consensus quality: 144233 bases at least Q30  
Consensus quality: 150066 bases at least Q20  
Estimated insert size: 132352; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-ff estimation  
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 74 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 5909: contig of 5909 bp in length  
\* 5910 6009: gap of unknown length  
\* 6010 10545: contig of 4536 bp in length  
\* 10546 10645: gap of unknown length  
\* 10646 14032: contig of 3387 bp in length  
\* 14033 14132: gap of unknown length  
\* 14133 17936: contig of 3804 bp in length  
\* 17937 18036: gap of unknown length  
\* 18037 22114: contig of 4078 bp in length  
\* 22115 26249: gap of unknown length  
\* 26250 31344: gap of unknown length  
\* 31345 31344: contig of 4995 bp in length

31345 31444: gap of unknown length  
\* 31445 35933: contig of 4489 bp in length  
\* 35934 36034: gap of unknown length  
\* 36034 40540: contig of 4507 bp in length  
\* 40541 40641: gap of unknown length  
\* 40641 43880: contig of 3239 bp in length  
\* 43880 43980: gap of unknown length  
\* 43980 48195: contig of 4215 bp in length  
\* 48195 48295: gap of unknown length  
\* 48295 52292: contig of 3998 bp in length  
\* 52292 52393: gap of unknown length  
\* 52393 55406: contig of 3014 bp in length  
\* 55406 55507: gap of unknown length  
\* 55507 57239: contig of 1733 bp in length  
\* 57239 57338: gap of unknown length  
\* 57338 60329: contig of 2989 bp in length  
\* 60329 60429: gap of unknown length  
\* 60429 63995: contig of 3567 bp in length  
\* 63995 64095: gap of unknown length  
\* 64095 66845: contig of 2750 bp in length  
\* 66845 66946: gap of unknown length  
\* 66946 70110: contig of 3165 bp in length  
\* 70110 70210: gap of unknown length  
\* 70210 72617: contig of 2407 bp in length  
\* 72617 72717: gap of unknown length  
\* 72717 75119: contig of 2802 bp in length  
\* 75119 75619: gap of unknown length  
\* 75619 77931: contig of 2312 bp in length  
\* 77931 78031: gap of unknown length  
\* 78031 80825: contig of 2794 bp in length  
\* 80825 80925: gap of unknown length  
\* 80925 83263: contig of 2338 bp in length  
\* 83263 83364: gap of unknown length  
\* 83364 86127: contig of 2763 bp in length  
\* 86127 86227: gap of unknown length  
\* 86227 89182: contig of 2956 bp in length  
\* 89182 89282: gap of unknown length  
\* 89282 90484: contig of 1202 bp in length  
\* 90484 90584: gap of unknown length  
\* 90584 92580: contig of 1996 bp in length  
\* 92580 92681: gap of unknown length  
\* 92681 94006: contig of 1326 bp in length  
\* 94006 94106: gap of unknown length  
\* 94106 96455: contig of 2349 bp in length  
\* 96455 96555: gap of unknown length  
\* 96555 98594: contig of 2039 bp in length  
\* 98594 98694: gap of unknown length  
\* 98694 100357: contig of 1663 bp in length  
\* 100357 100457: gap of unknown length  
\* 100457 102384: contig of 1927 bp in length  
\* 102384 102484: gap of unknown length  
\* 102484 104202: contig of 1718 bp in length  
\* 104202 104302: gap of unknown length  
\* 104302 105380: contig of 1078 bp in length  
\* 105380 105480: gap of unknown length  
\* 105480 107088: contig of 1608 bp in length  
\* 107088 107188: gap of unknown length  
\* 107188 108957: contig of 1769 bp in length  
\* 108957 109057: gap of unknown length  
\* 109057 110942: contig of 1865 bp in length  
\* 110942 111042: gap of unknown length  
\* 111042 113098: contig of 2056 bp in length  
\* 113098 113198: gap of unknown length  
\* 113198 115082: contig of 1884 bp in length  
\* 115082 115182: gap of unknown length  
\* 115182 116354: contig of 1172 bp in length  
\* 116354 116454: gap of unknown length  
\* 116454 118146: contig of 1692 bp in length  
\* 118146 118246: gap of unknown length  
\* 118246 119751: contig of 1505 bp in length  
\* 119751 119851: gap of unknown length  
\* 119851 121640: contig of 1789 bp in length  
\* 121640 121740: gap of unknown length



Query Match	20.3%;	Score 50.6;	DB 2;	Length 168213;
Best Local Similarity	34.6%;	Pred. No. 0.0091;		
Matches 71;	Conservative	0;	Mismatches 134;	Indels 0;
			Gaps	0;
*	121741	123612:	contig of 1872 bp in length	
*	123713	123712:	gap of unknown length	
*	123713	126065:	contig of 2353 bp in length	
*	126066	126165:	gap of unknown length	
*	126166	127275:	contig of 1110 bp in length	
*	127276	127375:	gap of unknown length	
*	127376	128763:	contig of 1388 bp in length	
*	128764	128863:	gap of unknown length	
*	130866	130267:	contig of 1404 bp in length	
*	130268	130367:	gap of unknown length	
*	130368	131975:	contig of 1608 bp in length	
*	131976	132075:	gap of unknown length	
*	132076	133942:	contig of 1867 bp in length	
*	133943	134042:	gap of unknown length	
*	134043	135320:	contig of 1278 bp in length	
*	135321	135420:	gap of unknown length	
*	135421	136607:	contig of 1187 bp in length	
*	136608	136707:	gap of unknown length	
*	136708	139099:	contig of 2352 bp in length	
*	139100	139199:	gap of unknown length	
*	139200	140405:	contig of 1206 bp in length	
*	140406	140505:	gap of unknown length	
*	140506	142113:	contig of 1608 bp in length	
*	142114	142213:	gap of unknown length	
*	142214	143820:	contig of 1607 bp in length	
*	143821	143920:	gap of unknown length	

· TITLE  
JOURNAL  
REFERENCEE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garera,N., Gill,R., Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Hawlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kravovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,Z., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Longueval,H., Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Matindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newson,N., Ngum,A., Ngyuen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S., Ogburn,M., Okonou,G., Oranuyn,N., Oviedo,R., Pace,A., Payton,B., Peety,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pul,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Roubakoum,T., Rolle,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Somaika,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tablor,P., Tameis,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vazquez,L., Vera,Y., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,C., Washington,C., Watlington,C., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 166280)

Worley,K.C.

Direct Submission

Submitted (25-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

on Dec 20, 2001>this sequence version replaced gi:17064266.

Dy	103	cctggaggtgcgaagcccaacataccatgcatcattcaaaaacaagaataaacagtca	222
Db	10574	NN	10633
Qy	223	tgcatatgtctatttttttatt 247         	
Db	10634	NNNNNNNNNNNNNTTCCTTTGCAAT 10658	

```

RESULT 10
AC103460
LOCUS AC103460 168280 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH30-30D24, *** SEQUENCING IN PROGRESS
ACCESSION AC103460
VERSION AC103460.2 GI:17973302
KEYWORDS HTG; HTGS_PHASEL.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

```

REFERENCE  
1 (pages 1 to 168280)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alshrooks, S.L., Ameratunge, H.C., Aze, J.R., Banks, T., Barbara, J., Benton, J., Bimoge, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowles, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Bukey, C., Butcher, P., Burrell, C., Burrell, K.L., Byrd, N.C., Carion, T.F., Cartch, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dean, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.U., Draper, H., Dugan-Rocha, S., Dumbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garera,N., Gill,R., Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Hawlak,P., Hayes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kralovic,J., Kreshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Liuweh,G., Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,G., Nguyen,N., Nickerson,E., Nwokankwo,S., Nqubu,N., Okonou,G., Oranuyn,N., Oviedo,R., Pace,A., Payton,B., Peety,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Roudbokan,T., Rolfe,M., Ruiz,S., Savery,G., Scheer,S., Scott,G., Shen,H., Shooshlari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabop,R., Tameris,A., Tameris,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vazquez,L., Vera,Y., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Unpublished  
Direct Submission  
2 (bases 1 to 168280)  
Worley,K.C.

Submitted (25-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:17064266.

Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: HGHP  
Center clone name: CH230-30P24  
Summary Statistics  
Assembly program: phrap; version 0.990329first call to findphraplast

Consensus quality: 135039 bases at least Q40  
Consensus quality: 141582 bases at least Q30  
Consensus quality: 147182 bases at least Q20  
Estimated insert size: 127529; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-gel estimation  
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
NOTE: This is a 'working draft' sequence. It currently consists of 74 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	7704:	contig of 7704 bp in length
7705	7805:	gap of unknown length
7805	13065:	contig of 5261 bp in length
13066	13165:	gap of unknown length
13166	19229:	contig of 6064 bp in length
19230	19329:	gap of unknown length
19330	24345:	contig of 5016 bp in length
24346	24445:	gap of unknown length





```

repeat_region /rpt_family="Alu"
10008..10237
/rpt_family="MIR"
repeat_region 10359..10717
/rpt_family="MaLR"
repeat_region 10877..10962
/rpt_family="MIR"
repeat_region 11116..11369
/rpt_family="L2"
repeat_region 11370..11672
/rpt_family="Alu"
repeat_region 11673..11747
/rpt_family="L2"
repeat_region 11845..12112
/rpt_family="ERVU"
repeat_region 12113..12342
/rpt_family="CR1"
repeat_region 12447..12561
/rpt_family="CR1"
repeat_region 13473..13601
/rpt_family="MaLR"
repeat_region 13629..13783
/rpt_family="MaLR"
repeat_region 16127..16428
/rpt_family="MER1_type"
repeat_region 16597..16794
/rpt_family="MIR"
repeat_region 16873..17103
/rpt_family="MIR"
17766..18565
/ncore="similar to EST BG213772 (NID:913735459)"
repeat_region 18021..18499
/rpt_family="MaLR"
repeat_region 18500..18872
/rpt_family="MaLR"
repeat_region 19401..19561
/rpt_family="MER1_type"
repeat_region 19821..19877
/rpt_family="Alu"
repeat_region 19881..20095
/rpt_family="Alu"
repeat_region 20108..20246
/rpt_family="MER1_type"
repeat_region 20347..20642
/rpt_family="Alu"
repeat_region 20738..20817
/rpt_family="Mariner"
repeat_region 21274..21322
/rpt_family="MIR"
repeat_region 21962..21985
/rpt_family="(TAAA)n"
repeat_region 22291..22383
/rpt_family="MER1_type"
repeat_region 22725..22821
/rpt_family="L2"
repeat_region 22837..23127
/rpt_family="Alu"
repeat_region 24736..24765
/rpt_family="(CA)n"
repeat_region 25411..25641
/rpt_family="MIR"
repeat_region 25685..25922
/rpt_family="Alu"
repeat_region 26039..26113
/rpt_family="Alu"
repeat_region 27823..28134
/rpt_family="L2"
repeat_region 28159..28333
/rpt_family="L2"
repeat_region 29197..29496
/rpt_family="Alu"
repeat_region 29621..29668
/rpt_family="AT_rich"

```

```

repeat_region 30581..30608
/rpt_family="AT_rich"
repeat_region 30805..30834
/rpt_family="AT_rich"
repeat_region 30967..31073
/rpt_family="MER1_type"
repeat_region 35404..35425
/rpt_family="AT_rich"
repeat_region 35508..35617
/rpt_family="CR1"
repeat_region 36725..36845
/rpt_family="MIR"
repeat_region 36931..37008
/rpt_family="MER1_type"

Query Match 15.6%; Score 38.8; DB 9; Length 176741;
Best Local Similarity 62.2%; Pred No. 2.1;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Oy 150 gatgtggaagaaagctgaggtgaagctcagcatccatgattacttaaaacagaa 209
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98525 GATGAGAGAGAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 98466
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy 210 aaaaagacatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 247
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98465 ATATATATATATATATATATATATATATATATATATATATATATATATAT 98428
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

## RESULT 13

AC023228

LOCUS AC023228 203194 bp DNA linear HTG 01-MAR-2000

DEFINITION Homo sapiens chromosome 2 clone RP11-615C5 map 2, WORKING DRAFT

AC023228

AC023228.2 GI:7139834

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 203194)

Blatten,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 2, clone RP11-615C5

Unpublished

2 (bases 1 to 203194)

Blatten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,

Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,

Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,

Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,

Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Landers,T., Lehoczyk,U., Levine,R., Lieu,C., Liu,G., Locke,K.,

Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,

Mcpheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,

Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rotman,D.,

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,

Zimmer,A. and Zody,M.

Direct Submission

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 1, 2000 this sequence version replaced gi:6957742.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: http://www-seq.wi.mit.edu

```

-----
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L6708
Center clone name: 615_C_5
-----
Summary Statistics
Sequencing vector: M13; M77815: 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 186899 bases at least Q40
Consensus quality: 195828 bases at least Q30
Consensus quality: 198475 bases at least Q20
Insert size: 185000; agarose-fp
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1491: contig of 1491 bp in length
*
1492 1591: gap of 100 bp
*
1592 2856: contig of 1265 bp in length
*
2857 2956: gap of 100 bp
*
2957 4454: contig of 1498 bp in length
*
4455 4554: gap of 100 bp
*
4555 5993: contig of 1439 bp in length
*
5994 6093: gap of 100 bp
*
6094 8196: contig of 2103 bp in length
*
8197 8296: gap of 100 bp
*
8297 9731: contig of 1435 bp in length
*
9732 9831: gap of 100 bp
*
9832 11084: contig of 1253 bp in length
*
11085 11184: gap of 100 bp
*
11185 12899: contig of 1715 bp in length
*
12900 12999: gap of 100 bp
*
13000 13527: contig of 528 bp in length
*
13528 13627: gap of 100 bp
*
13628 15114: contig of 1487 bp in length
*
15115 15214: gap of 100 bp
*
15215 16378: contig of 1164 bp in length
*
16379 16478: gap of 100 bp
*
16479 18131: contig of 1653 bp in length
*
18132 18231: gap of 100 bp
*
18232 20083: contig of 1852 bp in length
*
20084 20183: gap of 100 bp
*
20184 21117: contig of 934 bp in length
*
21118 21217: gap of 100 bp
*
21218 24091: contig of 2874 bp in length
*
24092 24191: gap of 100 bp
*
24192 28023: contig of 3388 bp in length
*
28030 28129: gap of 100 bp
*
28130 33133: contig of 5004 bp in length
*
33134 33233: gap of 100 bp
*
33234 37816: contig of 4583 bp in length
*
37817 37916: gap of 100 bp
*
37917 42224: contig of 4308 bp in length
*
42225 42324: gap of 100 bp
*
42325 48180: contig of 5856 bp in length
*
48181 48280: gap of 100 bp
*
48281 54554: contig of 6274 bp in length
*
54555 54654: gap of 100 bp
*
54655 60966: contig of 6312 bp in length
*
60967 61066: gap of 100 bp
*
61067 71645: contig of 10579 bp in length
*
71646 71745: gap of 100 bp
*
71746 81326: contig of 9581 bp in length
*
81327 81426: gap of 100 bp
*
81427 93995: contig of 12569 bp in length

```

```

FEATURES
source
*
93996 94095: gap of 100 bp
*
94096 103668: contig of 9573 bp in length
*
103669 103768: gap of 100 bp
*
103769 113879: contig of 10111 bp in length
*
113880 113979: gap of 100 bp
*
113980 128120: contig of 14141 bp in length
*
128121 128220: gap of 100 bp
*
128221 141049: contig of 12829 bp in length
*
141050 141149: gap of 100 bp
*
141150 155157: contig of 14008 bp in length
*
155158 155257: gap of 100 bp
*
155258 177981: contig of 22724 bp in length
*
177982 178081: gap of 100 bp
*
178082 203194: contig of 25113 bp in length.
Location/Qualifiers
1. 203194
/Organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-615C5"
/clone_lib="RPC1-11 Human Male BAC"
1. 1491
/note="assembly_fragment"
1592. 2856
/note="assembly_fragment"
2957. 4454
/note="assembly_fragment"
4555. 5993
/note="assembly_fragment"
6094. 8196
/note="assembly_fragment"
8297. 9731
/note="assembly_fragment"
9832. 11084
/note="assembly_fragment"
11185. 12899
/note="assembly_fragment"
13000. 13527
/note="assembly_fragment"
13628. 15114
/note="assembly_fragment"
15215. 16378
/note="assembly_fragment"
16479. 18131
/note="assembly_fragment"
18232. 20083
/note="assembly_fragment"
20184. 21117
/note="assembly_fragment"
24091. 28023
/note="assembly_fragment"
28129. 28130
/note="assembly_fragment"
33234. 37816
/note="assembly_fragment"
37917. 42224
/note="assembly_fragment"
42325. 48180
/note="assembly_fragment"
48281. 54554
/note="assembly_fragment"
54655. 60966
/note="assembly_fragment"
61067. 71645
/note="assembly_fragment"
71746. 81326
/note="assembly_fragment"

```

Query Match	15.48;	Score 38.4;	DB 9;	Length 119579;
-------------	--------	-------------	-------	----------------

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*      1 7790: contig of 7790 bp in length
*      7791 7890: gap of 100 bp
*      7891 19908: contig of 12018 bp in length
*      19909 20008: gap of 100 bp
*      20009 26287: contig of 6279 bp in length
*      26288 26387: gap of 100 bp
*      26388 46199: contig of 19812 bp in length
*      46200 46299: gap of 100 bp
*      46300 54017: contig of 7718 bp in length
*      54018 54117: gap of 100 bp
*      54118 64629: contig of 10512 bp in length

```

```

* 64630 64729: gap of 100 bp
* 64730 69964: contig of 5235 bp in length
* 69965 70064: gap of 100 bp
* 70065 79016: contig of 8952 bp in length
* 79017 79116: gap of 100 bp
* 79117 82147: contig of 3031 bp in length
* 82148 82247: gap of 100 bp
* 82248 92795: contig of 10548 bp in length.

```

## FEATURES

SOURCE

```

1. 92795
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="1"
  /map="p34.1-34.3"
  /clone="RPI-6283"
  /clone_1b="RPC1-1"
  1. 7790
  /note="assembly_fragment:00579
  fragment_chain:1
  clone_end:SP6
  vector_side:left"
  7891. 19908
  /note="assembly_fragment:00853
  fragment_chain:1"
  20089. 26287
  /note="assembly_fragment:00491
  fragment_chain:1"
  26388. 46199
  /note="assembly_fragment:00189
  fragment_chain:1"
  46300. 34017
  /note="assembly_fragment:00445
  fragment_chain:1"
  54118. 64629
  /note="assembly_fragment:00051
  fragment_chain:1"
  64730. 69964
  /note="assembly_fragment:00668
  fragment_chain:1"
  70065. 79016
  /note="assembly_fragment:00709
  fragment_chain:1"
  79117. 82147
  /note="assembly_fragment:00073
  fragment_chain:1"
  82248. 92795
  /note="assembly_fragment:00423
  clone_end:T7
  vector_side:right"

```

```

BASE COUNT 24608 a 19859 c 19994 g 27433 t 901 others
ORIGIN

```

```

Query Match 15.3%; Score 38.2; DB 2; Length 92795;
Best Local Similarity 78.0%; Pred. No. 3;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 189 gtatttacttaaaacagaaaagacatgtaggatatgtcatatttttttttatt 247
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 61779 GTCCTTCTTTAAAAAATAATATATATATATATATATATATATATATATATTTT 61837

```

```

Search completed: September 20, 2002, 05:26:44
Job time: 5124 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 08:30:59 ; Search time 1856.42 Seconds

(without alignments)  
2806.858 Million cell updates/sec

Title: US-09-757-781-21

Sequence: 1 cagcgtgtgtgtgcaaggaagc.....gtcatattttttttatg 249

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_da:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hg\_hum:\*  
31: em\_hg\_inv:\*  
32: em\_hg\_other:\*  
33: em\_hgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	249	100.0	2696	9	AK000761	AK000761 Homo sapi
2	249	100.0	5958	9	AF196185	AF196185 Homo sapi
3	249	100.0	22279	2	U82210	U82210 Homo sapien
4	249	100.0	106198	2	AL160409	AL160409 Human DNA
5	249	100.0	138688	2	AC025823	AC025823 Homo sapi
6	249	100.0	196374	2	AL591464	AL591464 Homo sapi
7	176	70.7	427	6	AX332939	AX332939 Sequence
8	70	28.1	5500	10	AB005549	AB005549 Rattus no
9	50.6	20.3	168213	2	AC103175	AC103175 Rattus no
10	40.8	16.4	168280	2	AC103460	AC103460 Rattus no
11	39	15.7	2305	3	S96842	S96842 sp96=spore-
12	38.8	15.6	176741	9	AC016732	AC016732 Homo sapi
13	38.6	15.5	203194	2	AC023228	AC023228 Homo sapi
14	38.4	15.4	119579	9	AC008425	AC008425 Homo sapi
15	38.2	15.3	92795	2	AL357035	AL357035 Homo sapi
16	38.2	15.3	198965	2	AC103542	AC103542 Rattus no
17	38	15.3	197239	2	AC094815	AC094815 Rattus no
18	37.6	15.1	162260	2	AC021236	AC021236 Homo sapi
19	37.4	15.0	55774	2	AC091303	AC091303 Homo sapi
20	37.2	14.9	76113	2	AC023453	AC023453 Homo sapi
21	37	14.9	125301	9	HS388N15	Z99571 Human DNA s
22	36.8	14.8	12029	3	AE001412	AE001412 Plasmodiu
23	36.8	14.8	185182	9	AC093423	AC093423 Homo sapi
24	36.6	14.7	39608	9	AC009004	AC009004 Homo sapi
25	36.6	14.7	74470	9	AC079404	AC079404 Homo sapi
26	36.6	14.7	79085	9	AC006221	AC006221 Homo sapi
27	36.6	14.7	164309	30	AC067834	AC067834 Homo sapi
28	36.6	14.7	166500	9	AC007785	AC007785 Homo sapi
29	36.6	14.7	169572	2	AC108219	AC108219 Homo sapi
30	36.6	14.7	191830	2	AC025660	AC025660 Homo sapi
31	36.6	14.7	193351	9	AC025656	AC025656 Homo sapi
32	36.6	14.7	205908	2	AL669903	AL669903 Mus muscu
33	36.4	14.6	34722	8	SPBC16C6	AL021767 S.pombe c
34	36.4	14.6	36629	2	AC107279	AC107279 Rattus no
35	36.4	14.6	186135	2	AL669927	AL669927 Mus muscu
36	36.2	14.5	7784	6	AX347326	AX347326 Sequence
37	36.2	14.5	110000	2	AP002753_2	Continuation (3 of
38	36.2	14.5	110000	10	AE008684_1	Continuation (2 of
39	36.2	14.5	131972	4	AC008794	AC008794 Homo sapi
40	36.2	14.5	132090	10	AC005402	AC005402 Mus muscu
41	36.2	14.5	151558	9	AC021766	AC021766 Homo sapi
42	36.2	14.5	155633	2	AC078795	AC078795 Homo sapi
43	36.2	14.5	156060	2	AC004153	AC004153 Plasmodiu
44	36.2	14.5	169546	2	AC004157	AC004157 Plasmodiu
45	36.2	14.5	169794	2	AC004688	AC004688 Plasmodiu

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE
1	AK000761	Homo sapiens cDNA FLJ20754 fls, clone HEP02246.	AK000761	1	GI:7021050	oligo capping; fls (full insert sequence).	Homo sapiens hepatoma cell_line:hepc2 cDNA to mRNA, clone_11b:HEP	1	Kawakami,T., Nouguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsunura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.	NEBO human cDNA sequencing project	Unpublished (2000)	2 (bases 1 to 2696)	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.	Direct Submission





Db	5582	GOMGGGGMGTCTAAATGGCCCTGTGAAATCCCGATGTGGAAAAAGCTGAGGTGAAGCTCAG	5641
QY	181	cataccatgtaattactcttaaaacagaaaaaagacatgtaatgatactattttt	240
Db	5642	CATACCAGTATTTACTTTAAAAACAGAAAAAAGACATGTATGATATGCTATTTTTT	5701
QY	241	tttatgtg 249	
Db	5702	TTTATGTGG 5710	
RESULT	3		
LOCUS	U82210	22279 bp DNA linear HTG 30-NOV-2001	
DEFINITION	Homo sapiens chromosome 10 clone CRI-JC2075 map 10p11.2, ***		
ACCESSION	U82210	SEQUENCING IN PROGRESS ***, 1 ordered pieces.	
VERSION	U82210.1	GI:3983510	
KEYWORDS	HTG; HTGS_PHASe2; HTGS_CANCELLED.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 22279)		
AUTHORS	Smith,D.R.		
TITLE	Sequencing of Human Chromosome 10		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 22279)		
AUTHORS	Du,L. and Smith,D.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-DEC-1996) Bioinformatics Division, Genome		
	Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA		
COMMENT	On Dec 9, 1998 this sequence version replaced gi:1773048.		

	Technology. Data may contain low quality sequence and BAC/Cosmid vector sequences.
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 1 contigs. Gaps between the contigs
	* are represented as runs of N. The order of the pieces
	* is believed to be correct as given, however the sizes
	* of the gaps between them are based on estimates that have
	* provided by the submittor.
	* This sequence will be replaced
	* by the finished sequence as soon as it is available and
	* the accession number will be preserved.
	1      22279; contig of 22279 bp in length.
	Location/Qualifiers
	1..22279
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="10"
	/map="10p11.2"
	/clone="CRI-JC2075"
BASE COUNT	5671 a 5238 c 5397 g 5959 t 14 others
ORIGIN	
Query Match	100.0%; Score 249; DB 2; Length 22279;
Best Local Similarity	100.0%; Pred. No. 1.be-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY 1	cagcggtggtgcgaagaagctactcctgcggycacatttaaggtytgtytgggtccg 60       Db 10975 CAGCGGTGTCGAAGAAGCTACTCCTGC CGCAATATTAGATGTGTGGGTCTCG 11034       OY 61 ggagatctcggtgcgcccaatcttcctcatcattyttctgaacacccgtlatitgaaccay 120       Db 11035 GGAGATCTCGGTGCGCTCCCATCTTCCATTAATTTGTGGAACANCTGTATTGAAACCATMG 11094       OY 121 gctggggtgcctaagtgcctgtgaatcccgatgttgaaaaagctggaagtgtaagctcaag 180       Db 11095 GCTGGG GTGCTAAAAGTGCTGTGAATCCCGAATGTGAAAAAACCTGGAGGTGAACCTCAG 11154 

QY	101	cataccagtgatatttactttaaaacagaaaaaagacatglatggaatcgtctattttt	240
Db	11155	cataccagtgatatttacttttataaaacagaaaaaaagacatgfatgcatfmgctctatttttt	11214
QY	241	ttttatgtg	249
Db	11215	ttttatgtg	11223
RESULT	4		
LOCUS	AL160409/c		
DEFINITION	Human DNA sequence from clone RP11-406D17 on chromosome 10,		
ACCESSION	AL160409		
VERSION	AL160409.12		
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Sycamore, N.		
JOURNAL	Direct Submission		
	Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,		
	CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone		
	requests: clonerequests@sanger.ac.uk		
COMMENT	On Jul 19, 2001 this sequence version replaced gi:14625535.		

```

FEATURES
source
    1. .106198
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="10"
        /clone="RP11-406D17"
        /clone_id="RP11-11.2"
    27. .160
        /note="MIR repeat: matches 3. .139 of consensus"
    989. .1039
        /note="MIR repeat: matches 93. .144 of consensus"
    1232. .1275
        /note="22 copies 2 mer ac 77% conserved"
repeat_region
repeat_region
repeat_region

```

```
repeat_region 1344. .1051 /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 4229. .4606 /note="MLTJ repeat: matches 1. .403 of consensus"
repeat_region 4648. .4815 /note="FRAM repeat: matches 4. .171 of consensus"
repeat_region 5215. .5516 /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region 6161. .6216 /note="MIR repeat: matches 47. .104 of consensus"
repeat_region 6217. .6527 /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region 6528. .6637 /note="MIR repeat: matches 104. .205 of consensus"
repeat_region 6646. .6950 /note="AluJb repeat: matches 1. .304 of consensus"
repeat_region 8061. .8354 /note="AluJb repeat: matches 1. .289 of consensus"
repeat_region 8443. .8602 /note="AluJb/FRAM repeat: matches 152. .308 of consensus"
repeat_region 10005. .10072 /note="L2 repeat: matches 2673. .2741 of consensus"
repeat_region 10542. .10849 /note="MER57B repeat: matches 43. .386 of consensus"
repeat_region 10851. .10975 /note="FLAM.C repeat: matches 15. .133 of consensus"
repeat_region 11451. .11758 /note="AluJb repeat: matches 1. .312 of consensus"
repeat_region 12992. .14459 /note="PTR5 repeat: matches 504. .1874 of consensus"
repeat_region 14460. .14550 /note="MER61E repeat: matches 377. .466 of consensus"
repeat_region 14551. .14771 /note="PTR5 repeat: matches 1874. .8522 of consensus"
repeat_region 14772. .14923 /note="AluSx repeat: matches 3. .154 of consensus"
repeat_region 14930. .15218 /note="AluJb repeat: matches 1. .293 of consensus"
repeat_region 15223. .15519 /note="AluSg repeat: matches 1. .299 of consensus"
repeat_region 15661. .15818 /note="MER5A repeat: matches 34. .189 of consensus"
repeat_region 15868. .15919 /note="26 copies 2 mer aa 73% conserved"
repeat_region 16133. .16443 /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 16599. .16707 /note="L1MC4 repeat: matches 7728. .7834 of consensus"
repeat_region 16712. .16787 /note="38 copies 2 mer ta 76% conserved"
repeat_region 16825. .17092 /note="AluJb repeat: matches 1. .292 of consensus"
repeat_region 17094. .17254 /note="AluJb/FRAM repeat: matches 134. .295 of consensus"
repeat_region 17579. .17889 /note="AluSg repeat: matches 1. .306 of consensus"
repeat_region 18807. .19086 /note="MLTJ repeat: matches 3. .238 of consensus"
repeat_region 19087. .19391 /note="AluSg repeat: matches 1. .310 of consensus"
repeat_region 19392. .19542 /note="MLTJ repeat: matches 238. .501 of consensus"
repeat_region 19608. .19904 /note="AluJb repeat: matches 1. .295 of consensus"
repeat_region 20333. .20642 /note="AluSx repeat: matches 1. .310 of consensus"
repeat_region 20670. .20956 /note="AluJb repeat: matches 1. .286 of consensus"
repeat_region 21297. .21601 /note="L2 repeat: matches 2414. .2696 of consensus"
repeat_region 21857. .22175 /note="AluJb repeat: matches 1. .311 of consensus"
repeat_region 22502. .22739 /note="AluSx repeat: matches 1. .239 of consensus"
repeat_region 23135. .23915 /note="L1PA10 repeat: matches 5369. .6158 of consensus"
repeat_region 27500. .27629 /note="L1MA3 repeat: matches 6171. .6304 of consensus"
repeat_region 28661. .28984 /note="AluSg repeat: matches 1. .312 of consensus"
repeat_region 30001. .30132 /note="66 copies 2 mer aa 62% conserved"
repeat_region 30233. .30328 /note="48 copies 2 mer ta 65% conserved"
repeat_region 30365. .30644 /note="L1MC4 repeat: matches 6633. .6940 of consensus"
repeat_region 30645. .31062 /note="MSTB repeat: matches 1. .424 of consensus"
repeat_region 31063. .31122 /note="L1MC4 repeat: matches 6940. .6995 of consensus"
repeat_region 31133. .31547 /note="L1MC4 repeat: matches 7167. .7597 of consensus"
repeat_region 31648. .32167 /note="L1PA6 repeat: matches 5612. .6131 of consensus"
repeat_region 32275. .32552 /note="AluJb repeat: matches 5. .282 of consensus"
repeat_region 32592. .32686 /note="L1MC4 repeat: matches 7791. .7972 of consensus"
repeat_region 33154. .33272 /note="MIR repeat: matches 80. .185 of consensus"
repeat_region 33283. .33415 /note="AluJb repeat: matches 14. .153 of consensus"
repeat_region 33416. .34701 /note="PTR5 repeat: matches 172. .2212 of consensus"
repeat_region 34713. .34840 /note="AluJb/FRAM repeat: matches 178. .302 of consensus"
repeat_region 35365. .35654 /note="AluJb repeat: matches 1. .288 of consensus"
repeat_region 36067. .36155 /note="L2 repeat: matches 2622. .2710 of consensus"
repeat_region 36582. .36804 /note="MER30 repeat: matches 1. .230 of consensus"
repeat_region 37547. .37852 /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region 37894. .38050 /note="MER5A repeat: matches 9. .188 of consensus"
repeat_region 38447. .38744 /note="AluSx repeat: matches 5. .302 of consensus"
repeat_region 39063. .39193 /note="FLAM.C repeat: matches 1. .131 of consensus"
repeat_region 39510. .39673 /note="MER5A repeat: matches 4. .189 of consensus"
repeat_region 40932. .41102 /note="L1PA7 repeat: matches 1497. .1667 of consensus"
repeat_region 41103. .41726 /note="L1PA7 repeat: matches 5515. .6138 of consensus"
repeat_region 43386. .43747 /note="MER1B repeat: matches 1. .364 of consensus"
repeat_region 44459. .44771 /note="AluSg repeat: matches 3. .306 of consensus"
repeat_region 44991. .45302 /note="AluJb repeat: matches 1. .304 of consensus"
repeat_region 45536. .51640 /note="L1PA7 repeat: matches 11. .6145 of consensus"
repeat_region 51663. .51964 /note="AluJb repeat: matches 1. .308 of consensus"
repeat_region 51966. .52183 /note="109 copies 2 mer aa 69% conserved"

Query Match 100.0%; Score 249; DB 9; Length 106198;
Best Local Similarity 100.0%; Pred. No. 2e-60;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 cagcggtgtgtgcaggaagctactcgtcgtcaagtttagagtggtgtggtctcg 60
DB 86101 CACGCGTGTGTGCGAGGACGCTACCTCGCGTCAGTATAGAGTGTGTGTGCGTCG 86042
```

Qy	61	ggagactcggtaggcgcccaatctctctcaatgctcttgacaacccgtaattgnaaacatg	120
Db	8804.1	GGGAGCTCGGTGGGCTCCCATCTCTCTTCAATTGTTCGAACATCCTGATTGTGAAACCATG	87982
Qy	121	gctgggggtgctaaagtgtcgctgtaatcccgatgtygnaaaaagctgtaggtygnaagctcag	180
Db	8798.1	GCTGGGGGTGCTTAAATGCGCTGTGATCCGAGTGTGAAAAAGCTGGAGGTGAAGCTCAG	87922
Qy	181	cataccatgtaattctcctttaaacaagaacaaagacatgtaatgatatgctatttttt	240
Db	8792.1	CATACCAGTATATTACTTTAAAAACGAAAAAAGACATGATGCAATATGCTATTTT	87862
Qy	241	tttatattg	249
Db	8786.1	TTTTATG	87853

	RESULT	5		
AC025823/c	LOCUS	AC025823	138688 bp	DNA linear HTG_24-JAN-2002
DEFINITION	Homo sapiens chromosome 10 clone RP11-60H16,			WORKING DRAFT SEQUENCE, 7 unordered pieces.
ACCESSION	AC025823			
VERSION	AC025823.6	GI:13811888		
KEYWORDS	HTG; HTGS__PHASE1;	HTGS__DRAFT;	HTGS__CANCELLED.	
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 138688)			
AUTHORS	Smith,D.R.			
TITLE	Genome Therapeutics Corporation Sequencing Center: Human Genome Sequence Data Unpublished			
JOURNAL	2 (bases 1 to 138688) Smith,D.R.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (16-MAR-2000) Genome Therapeutics Corporation, 100 Beaver Street, Wellesham, MA 02453, USA			
JOURNML	On Apr 27, 2001 this sequence version replaced gi:9887686.			
COMMENT				

```

Quality coverage: 6.3x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
* be preserved.
*
* 1
* 1096 1095: contig of 1095 bp in length
* 1196 1195: gap of unknown length
* 2203 2203: contig of 1008 bp in length
* 2304 2303: gap of unknown length
* 9462 9462: contig of 7159 bp in length
* 9463 9562: gap of unknown length
* 9563 25807: contig of 16245 bp in length

```

FEATURES.	location/Qualifiers
SOURCE	1..138688
*	25808 25607: gap of unknown length
*	25908 39871: contig of 13864 bp in length
*	39872 39971: gap of unknown length
*	39972 68325: contig of 28354 bp in length
*	68326 68425: gap of unknown length
*	68426 138688: contig of 70263 bp in length

BASE ORIGIN	COUNT	41112 a	30000 c	29755 g	37221 t	600 others
misc_feature	1.	.1095				
misc_feature	/note="assembly_name:Contig14"	1196.	.2203			
misc_feature	/note="assembly_name:Contig15"	2304.	.9462			
misc_feature	/note="assembly_name:Contig15"	clone_end:17"				
misc_feature	9363.	.25807				
misc_feature	/note="assembly_name:Contig56"	clone_end:5P6"				
misc_feature	25908.	.39871				
misc_feature	/note="assembly_name:Contig57"	39972.	.68325			
misc_feature	/note="assembly_name:Contig58"	68426.	.138688			
misc_feature	/note="assembly_name:Contig59"					

Query Match	100.0%;	Score 249;	DB 2;	Length 138688;
Best Local Similarity	100.0%;	Pred. No. 2.1e-60;		
Matches 249;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 caacggtgtgtgcaagaaagctcaactctcgatcagtaattagaagtgtgtgtgtgctctcg 60  
|||  
Db 134742 CAGCGGTGGTGGCAGGAAGCTCACATCTCGCGTCAGATTAGAGTGTGTGTGTGCTCG 134683

Qy 61 ggagatctcggtggtcccatcttccctcatgttctcyaacatccgtgtttgtaaacatg 120  
|||||  
Db 134682 GGGATCTCGGTGGCTCCCATCTTCCATTCATTGTCGACATCCTGTATTGTAAACCATG 134623

Qy 121 gctggggtcctaagttgcctgtgaatcccgatbtgcaaaaagctggaagtgaagctcag 180  
|||||  
Db 134622 GCTGGGGTCTAAAGTGCCTGTGAATCCCGATGTGCAAAAAGCTGGAGGTGAAGCTCAG 134563

Qy 181 cataccatctattactcttaaaacagaaaaagacatgcatgcatgctctcttttc 240  
|||||  
Db 134562 CATACCATGTATTACTTTAAACAGAAAAAAGACATGTATGTCTATTTTTTT 134503

```
QY      241  ttattatgg  249
          |||||
Db 134502 TTTTATTGG 134494
```

RESULT	6
AL591464	
LOCUS	AL591464 196374 bp DNA linear HTG 19-DEC-2001
DEFINITION	Homo sapiens chromosome 10 clone RP3-302D10, *** SEQUENCING IN PROGRESS ***, 14 unordered pieces.
ACCESSION	AL591464
VERSION	AL591464.2 GI:17973978
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (sites)
TITLE	Burton, J.
	Direct Submission

JOURNAL

COMMENT

Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Dec 20, 2001 this sequence version replaced g1.14141513.

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

Project Information

Center project name: hb302D10

Summary Statistics

Sequencing program: XGAP4; version 4.5

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 189117 bases at least Q40

Consensus quality: 192136 bases at least Q30

Consensus quality: 193772 bases at least Q20

Insert size: 195074; sum-of-coverage

Quality coverage: 4.57x in Q20 bases; sum-of-coverage

Quality coverage: 4.33x in Q20 bases; agarose-tp

\* NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 40923: contig of 40923 bp in length  
40924 41023: gap of 100 bp  
41024 52424: contig of 11401 bp in length  
52425 52524: gap of 100 bp  
52525 56668: contig of 4144 bp in length  
56669 56768: gap of 100 bp  
56769 66064: contig of 9296 bp in length  
66065 66164: gap of 100 bp  
66165 74102: contig of 7938 bp in length  
74103 74202: gap of 100 bp  
74203 118025: contig of 43823 bp in length  
118026 118125: gap of 100 bp  
118126 120523: contig of 2398 bp in length  
120524 120623: gap of 100 bp  
120624 134346: contig of 13723 bp in length  
134347 134446: gap of 100 bp  
134447 143138: contig of 8692 bp in length  
143139 143238: gap of 100 bp  
143239 151820: contig of 8582 bp in length  
151821 151920: gap of 100 bp  
151921 179327: contig of 27407 bp in length  
179328 179427: gap of 100 bp  
179428 187106: contig of 7679 bp in length  
187107 187206: gap of 100 bp  
187207 193223: contig of 6017 bp in length  
193224 193323: gap of 100 bp  
193324 196374: contig of 3051 bp in length.

FEATURES

SOURCE

1. 196374  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP13-302D10"  
/clone\_id="RPC1-13.2"  
1. 40923  
/note="assembly\_fragment:00704  
fragment\_chain:1"  
41024. 52424  
/note="assembly\_fragment:02076  
fragment\_chain:1"  
52525. 56668  
/note="assembly\_fragment:02444

fragment\_chain:1"  
56769. 66064  
/note="assembly\_fragment:00688  
fragment\_chain:1"

misc-feature  
66165. 74102  
/note="assembly\_fragment:01900  
fragment\_chain:1"

misc-feature  
74203. 118025  
/note="assembly\_fragment:02289  
fragment\_chain:1"

misc-feature  
118126. 120523  
/note="assembly\_fragment:01018  
fragment\_chain:2"

misc-feature  
120624. 134346  
/note="assembly\_fragment:01079  
fragment\_chain:2"

misc-feature  
134447. 143138  
/note="assembly\_fragment:00344  
fragment\_chain:2"

misc-feature  
143239. 151820  
/note="assembly\_fragment:01621  
fragment\_chain:2"

misc-feature  
151921. 179327  
/note="assembly\_fragment:01247  
fragment\_chain:3"

misc-feature  
179428. 187106  
/note="assembly\_fragment:01160  
fragment\_chain:3"

misc-feature  
187207. 193223  
/note="assembly\_fragment:01995  
fragment\_chain:3"

misc-feature  
193324. 196374  
/note="assembly\_fragment:02022  
clone\_end:SP6  
vector\_side:right"

misc-feature  
BASE COUNT 57987 a 41071 c 41893 g 54114 t 1309 others

misc-feature  
ORIGIN

Query Match 100.0%; Score 249; DB 2; Length 196374;  
Best Local Similarity 100.0%; Pred. No. 2.1e-60;  
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcggtgtgtgcaagaaagctcctcgtcgtatgaatgaatgtgtgtgtgtgtcgc 60  
|||||  
DB 191512 CAGCGGTGTGTGCAAGAAAGCTCCTCGTCAGATTAGAGTGTGTGGTCTCG 191571

QY 61 gggatcgtgtgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 120  
|||||  
DB 191572 GGGATCTGCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 191631

QY 121 gctgtgtgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 180  
|||||  
DB 191632 GCTGTGTGTGTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 191691

QY 181 catcacatgtattactttaaataaagaaagaaagaaagaaagaaagaaagaaagaaagaa 240  
|||||  
DB 191692 CATCACATGTATTACTTTAAATAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 191751

QY 241 tttttttttt 249  
|||||  
DB 191752 TTTTATTGCG 191760

RESULT 7  
AX332939/c 427 bp DNA linear PAT 09-JAN-2002  
LOCUS AX332939 Sequence 3448 from Patent WO0194629.  
DEFINITION AX332939  
ACCESSION AX332939  
VERSION AX332939.1 GI:18123573  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

[illegible]

Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvach, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtenberg, O., Lien, C., Liu, J., Liu, W., Ma, U., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, J., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Oguni, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, R., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D., Weinstein, G., and GIBBS, R.

Unpublished  
2 (bases 1 to 168213)  
Worley, K.C.

Direct Submission  
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Dec 21, 2001 this sequence version replaced gi:17062822.

Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

Project Information  
Center project name: GJXC  
Center clone name: CH230-98H12

Summary Statistics  
-----  
Assembly program: Phrap; version 0.990329First call to findPhrapList

Consensus quality: 136914 bases at least Q40  
Consensus quality: 144233 bases at least Q30  
Consensus quality: 150066 bases at least Q20  
Estimated insert size: 132352; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-ff estimation  
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 74 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 5909: contig of 5909 bp in length  
\* 5910 6009: gap of unknown length  
\* 6010 10545: contig of 4536 bp in length  
\* 10546 10645: gap of unknown length  
\* 10646 14032: contig of 3387 bp in length  
\* 14033 14132: gap of unknown length  
\* 14133 17936: contig of 3804 bp in length  
\* 17937 18036: gap of unknown length  
\* 18037 22114: contig of 4078 bp in length  
\* 22115 26249: gap of unknown length  
\* 26250 31344: gap of unknown length  
\* 31344: contig of 4995 bp in length  
\* 26350

31345 31444: gap of unknown length  
\* 31445 35933: contig of 4489 bp in length  
\* 35934 36033: gap of unknown length  
\* 36034 40540: contig of 4507 bp in length  
\* 40541 40640: gap of unknown length  
\* 40641 43879: contig of 3239 bp in length  
\* 43880 43979: gap of unknown length  
\* 43980 48194: contig of 4215 bp in length  
\* 48195 48294: gap of unknown length  
\* 48295 52292: contig of 3998 bp in length  
\* 52293 52392: gap of unknown length  
\* 52393 55406: contig of 3014 bp in length  
\* 55407 55506: gap of unknown length  
\* 55507 57229: contig of 1733 bp in length  
\* 57230 57339: gap of unknown length  
\* 57340 60328: contig of 2989 bp in length  
\* 60329 60429: gap of unknown length  
\* 60429 63995: contig of 3567 bp in length  
\* 63996 64095: gap of unknown length  
\* 64096 66845: contig of 2750 bp in length  
\* 66846 66945: gap of unknown length  
\* 66946 70110: contig of 3165 bp in length  
\* 70111 70210: gap of unknown length  
\* 70210 72617: contig of 2407 bp in length  
\* 72618 72717: gap of unknown length  
\* 72717 75519: contig of 2802 bp in length  
\* 75520 75619: gap of unknown length  
\* 75620 77931: contig of 2312 bp in length  
\* 77932 78031: gap of unknown length  
\* 78032 80825: contig of 2794 bp in length  
\* 80826 80925: gap of unknown length  
\* 80926 83263: contig of 2338 bp in length  
\* 83264 83363: gap of unknown length  
\* 83364 86126: contig of 2763 bp in length  
\* 86127 86226: gap of unknown length  
\* 86227 89182: contig of 2956 bp in length  
\* 89183 89282: gap of unknown length  
\* 89283 90484: contig of 1202 bp in length  
\* 90485 90584: gap of unknown length  
\* 90585 92580: contig of 1996 bp in length  
\* 92581 92680: gap of unknown length  
\* 92681 94006: contig of 1326 bp in length  
\* 94007 94106: gap of unknown length  
\* 94107 96455: contig of 2349 bp in length  
\* 96456 96555: gap of unknown length  
\* 96556 98594: contig of 2039 bp in length  
\* 98595 98694: gap of unknown length  
\* 98695 100357: contig of 1663 bp in length  
\* 100358 100457: gap of unknown length  
\* 100458 102384: gap of unknown length  
\* 102385 102484: gap of unknown length  
\* 102485 104202: contig of 1718 bp in length  
\* 104203 104302: gap of unknown length  
\* 104303 105380: contig of 1078 bp in length  
\* 105381 105480: gap of unknown length  
\* 105481 107088: contig of 1608 bp in length  
\* 107089 107188: gap of unknown length  
\* 107189 108957: contig of 1769 bp in length  
\* 108958 109057: gap of unknown length  
\* 109059 110942: contig of 1885 bp in length  
\* 110943 111042: gap of unknown length  
\* 111043 113098: contig of 2056 bp in length  
\* 113099 113198: gap of unknown length  
\* 113199 115082: contig of 1884 bp in length  
\* 115083 115182: gap of unknown length  
\* 115183 116354: contig of 1172 bp in length  
\* 116355 116454: gap of unknown length  
\* 116455 118146: contig of 1692 bp in length  
\* 118147 118246: gap of unknown length  
\* 118247 119751: contig of 1505 bp in length  
\* 119752 119851: gap of unknown length  
\* 119852 121640: contig of 1789 bp in length  
\* 121641 121740: gap of unknown length

[illegible]

Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Gatz, N., Gill, R., Gorrell, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hui, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, O., King, L., Korvah, J., Kovar, K., Krivov, J., Kuresh, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, Z., Lichtarge, O., Ileanu, C., Liu, J., Liu, W., Lonsaged, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nockentwo, S., Oguh, M., Okwunolu, G., Oragunye, N., Oyedero, R., Pace, A., Payton, B., Peery, J., Peter, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojudoan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 168280)  
Worley, K. C.

Direct Submission  
Submitted (23-Nov-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:17064266.

-----  
Genome Center  
-----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
-----  
Project Information  
-----  
Center project name: GHQP  
Center clone name: CH230-30D24  
-----  
Summary Statistics  
-----  
Assembly program: Phrap; version 0.990329First call to findhaplulist  
Consensus quality: 135039 bases at least Q40  
Consensus quality: 141582 bases at least Q30  
Consensus quality: 147182 bases at least Q20  
Estimated insert size: 127529; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agatools-fp estimation  
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

-----  
NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
NOTE: This is a 'working draft' sequence. It currently  
consists of 74 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 7704: contig of 7704 bp in length  
7705 7804: gap of unknown length  
7805 13065: contig of 5261 bp in length  
13066 13165: gap of unknown length  
13166 19229: contig of 6064 bp in length  
19230 19329: gap of unknown length  
19330 24345: contig of 5016 bp in length  
24346 24445: gap of unknown length



*	24466	30418:	contlig of 5973	bp in length
*	30419	30518:	gap of unknown length	
*	30519	35208:	contlig of 4690	bp in length
*	35209	35308:	gap of unknown length	
*	35309	39825:	contlig of 4517	bp in length
*	39826	39825:	gap of unknown length	
*	39926	45203:	contlig of 5278	bp in length
*	45204	45303:	gap of unknown length	
*	45304	46089:	contlig of 2786	bp in length
*	46090	46819:	gap of unknown length	
*	46819	53727:	contlig of 5538	bp in length
*	53728	53827:	gap of unknown length	
*	53828	57721:	contlig of 3894	bp in length
*	57722	57821:	gap of unknown length	
*	57822	60628:	contlig of 2807	bp in length
*	60629	60728:	gap of unknown length	
*	60729	60728:	gap of unknown length	
*	62775	65874:	gap of unknown length	
*	62875	65537:	contlig of 2663	bp in length
*	65538	65637:	gap of unknown length	
*	65638	67618:	contlig of 1981	bp in length
*	67619	67718:	gap of unknown length	
*	67719	71329:	contlig of 3511	bp in length
*	71230	71329:	gap of unknown length	
*	71330	73037:	contlig of 1708	bp in length
*	73038	73137:	gap of unknown length	
*	73138	76980:	contlig of 3843	bp in length
*	76981	77080:	gap of unknown length	
*	77081	77961:	contlig of 2881	bp in length
*	79962	80061:	gap of unknown length	
*	80062	82030:	contlig of 1969	bp in length
*	82031	82133:	gap of unknown length	
*	82131	84096:	contlig of 1966	bp in length
*	84097	84196:	gap of unknown length	
*	84197	85254:	contlig of 1058	bp in length
*	85255	85354:	gap of unknown length	
*	85355	87143:	contlig of 1789	bp in length
*	87144	87243:	gap of unknown length	
*	87244	89380:	contlig of 2137	bp in length
*	89381	89480:	gap of unknown length	
*	89481	91425:	contlig of 1945	bp in length
*	91426	91525:	gap of unknown length	
*	91526	93038:	contlig of 1513	bp in length
*	93039	93138:	gap of unknown length	
*	93139	95251:	contlig of 2113	bp in length
*	95252	95351:	gap of unknown length	
*	95352	96994:	contlig of 1643	bp in length
*	96995	97094:	gap of unknown length	
*	97095	99061:	contlig of 1967	bp in length
*	99062	99161:	gap of unknown length	
*	99162	101600:	contlig of 2439	bp in length
*	101601	101700:	gap of unknown length	
*	101701	104253:	contlig of 2553	bp in length
*	104254	104353:	gap of unknown length	
*	104354	106510:	contlig of 2157	bp in length
*	106511	106610:	gap of unknown length	
*	106611	106830:	contlig of 1820	bp in length
*	1068431	108530:	gap of unknown length	
*	108531	110001:	contlig of 1471	bp in length
*	110002	110101:	gap of unknown length	
*	110102	111861:	contlig of 1760	bp in length
*	111862	111961:	gap of unknown length	
*	111962	113383:	contlig of 1322	bp in length
*	113384	113383:	gap of unknown length	
*	113384	116326:	contlig of 2943	bp in length
*	116327	116426:	gap of unknown length	
*	116427	117867:	contlig of 1441	bp in length
*	117868	117967:	gap of unknown length	
*	117968	119282:	contlig of 1315	bp in length
*	119283	119382:	gap of unknown length	
*	119383	120322:	contlig of 1140	bp in length
*	120323	120622:	gap of unknown length	
*	120653	121770:	contlig of 1148	bp in length

Query Match	Best Local Similarity	16.4%;	Score 40.8;	DB 2;	Length 168280;
Matches	99; Conservative	16.4%;	Pred. No. 0.57;	Mismatches 97;	Indels 0; Gaps 0;
Qy 54	ggtctgggagatcgcgtggtccatcttcctgattgctgaacatcctgattgta 113				
Db 127130	GGTCATMGAGCTACCTCGGGCAATCTTTTCTTTCTTCTCATCTCCCTGTGGAA 127189				
Qy 114	aacctgctggggcgctcaagtcgctgtgaatcccgatgtgaaaaagctgaggtgaa 173				
Db 127190	TGCTATGCAAAAGGAAAAACCCCTATTAGTACAGATGTATGGAAAAACCAAGGTCAA 127249				
Qy 174	agctcagcatcacatgattacttctaataaacagaaaaaagacatgtatgatatgct 233				
Db 127250	CACCTCTCTCTCCCTTTCTCTCTATTTACTATGATTAATAATTAATAAGGTGGAGATGTCA 127309				
Qy 234	attttttttattatg 249				
Db 127310	TGGTGTGTGTGGG 127325				
RESULT 11					
LOCUS	S96842	2305 bp	DNA	linear	INV 07-MAY-1993
DEFINITION	SP96-spore-coat protein [5' region] [Dictyostelium discoideum, prespore cells, Genomic, 2305 nt].				
ACCESSION	S96842				
VERSION	S96842.1	GI:248952			
KEYWORDS					
SOURCE	Dictyostelium discoideum prespore cells.				
ORGANISM	Dictyostelium discoideum				
REFERENCE	Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.				
AUTHORS	1 (bases 1 to 2305)				
TITLE	Takemuchi, I.				
JOURNAL	Protein binding and DNase-I-hypersensitive sites in the cis-acting regulatory region of the spore-coat SP96 gene of Dictyostelium				
MEDLINE	Mech. Dev. 36 (3), 105-115 (1992)				
REMARK	92239372				
	Genbank staff at the National Library of Medicine created this entry [NCBI gistsq.96842] from the original journal article.				



repeat_region	30581..30608	/rpt_family="AT-rich"
repeat_region	30805..30834	/rpt_family="AT-rich"
repeat_region	30967..31073	/rpt_family="AT-rich"
repeat_region	35404..35425	/rpt_family="MER1_type"
repeat_region	35508..35617	/rpt_family="AT-rich"
repeat_region	36725..36845	/rpt_family="CR1"
repeat_region	36931..37008	/rpt_family="MIR"
repeat_region	36931..37008	/rpt_family="MER1_type"
Query Match	15.6%; Score 38.8; DB 9; Length 176741;	
Best Local Similarity	62.2%; Pred. No. 2.1;	
Matches	61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;	
Qy	150 gatgtggaanaagctcgtgaggtggaagcgcacatcaccatgatacttacttaaaacagaa 209	
Db	98525 GATGAGAGAGCGAGCGAGATGACATCTGGCGTGATTTGATTAATATATATATATATAT 98466	
Qy	210 aaaaacacatgatacgatactctatcttttttttatt 247	
Db	98465 AT 98428	
RESULT 13		
AC023228		
LOCUS	AC023228 203194 bp DNA linear HTG 01-MAR-2000	
DEFINITION	Homo sapiens chromosome 2 clone RP11-615C5 map 2, WORKING DRAFT	
SEQUENCE	32 unordered pieces.	
AC023228		
VERSION	AC023228.2 GI:7139834	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
TITLE	1 (bases 1 to 203194)	
JOURNAL	Bitren, B., Linton, L., Nusbaum, C. and Lander, E.	
REFERENCE	Homo sapiens chromosome 2, clone RP11-615C5	
AUTHORS	Unpublished	
2 (bases 1 to 203194)		
Bitren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,		
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,		
Boguslavsky, L., Boukhalter, B., Brown, A., Burrell, G., Castle, A.,		
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,		
DeRellano, K., Dewar, K., Domlo, M., Doyle, M., Feneater, J.,		
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,		
Gardyna, S., Grant, G., Hagos, B., Heald, A., Horton, L.,		
Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,		
Landers, T., Lechoczky, J., Levine, R., Lien, C., Liu, G., Locke, K.,		
Macdonald, P., Margolis, N., McEwan, P., McGurk, A., McKernan, K.,		
McHeeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,		
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,		
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,		
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,		
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,		
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,		
Zimmer, A. and Zody, M.		
Direct Submission		
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome		
Research, 320 Charles Street, Cambridge, MA 02141, USA		
On Mar 1, 2000 this sequence version replaced gi:6957742.		
All repeats were identified using RepeatMasker:		
Smit, A.F.A. & Green, P. (1996-1997)		
http://ftp.genome.washington.edu/RM/RepeatMasker.html		
----- Genome Center		
Center: Whitehead Institute/ MIT Center for Genome Research		
Web site: http://www-seq.wi.mit.edu		

\* 81427 93995: contig of 12569 bp in length

/note="assembly\_fragment"

msc_feature	71746.	.81326
-------------	--------	--------



```

* 64630 64729: gap of 100 bp
* 64730 69964: contig of 5235 bp in length
* 69965 70064: gap of 100 bp
* 70065 79016: contig of 8952 bp in length
* 79017 79116: gap of 100 bp
* 79117 82147: contig of 3031 bp in length
* 82148 82247: gap of 100 bp
* 82248 92795: contig of 10548 bp in length.

```

```

FEATURES
    source
        1. 92795
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /map="p34.1-34.3"
            /clone="RP1-62B3"
            /clone_lib="RPCI-1"
            1. 7790
                /note="assembly_fragment:00579
                    fragment_chain:1
                    clone_end:SP6
                    vector_side:left"
                7891. 19908
                    /note="assembly_fragment:00853
                        fragment_chain:1"
                2009. 26287
                    /note="assembly_fragment:00491
                        fragment_chain:1"
                26388. 46199
                    /note="assembly_fragment:00189
                        fragment_chain:1"
                46300. 54017
                    /note="assembly_fragment:00445
                        fragment_chain:1"
                5418. 64629
                    /note="assembly_fragment:00051
                        fragment_chain:1"
                64730. 69964
                    /note="assembly_fragment:00668
                        fragment_chain:1"
                70065. 79016
                    /note="assembly_fragment:00709
                        fragment_chain:1"
                79117. 82147
                    /note="assembly_fragment:00073
                        fragment_chain:1"
                82248. 92795
                    /note="assembly_fragment:00423
                        clone_end:T7
                        vector_side:right"

```

```

BASE COUNT 24608 a 19859 c 19994 g 27433 t 901 others
ORIGIN

```

```

Query Match 15.38; Score 38.2; DB 2; Length 92795;
Best Local Similarity 78.08; Pred. No. 3;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

OY 189 gttattactttaaaacagaaaagacagatgagatgctattttttttatt 247
Db 61779 GTCTTTCTTTTAAAAAATAATATATATATATATATATATTTT 61837

```

```

Search completed: September 21, 2002, 10:14:35
Job time: 6216 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 08:32:09 : Search time 180.75 Seconds  
(without alignments)  
2365.210 Million cell updates/sec

Title: US-09-757-781-21

Perfect score: 249

Sequence: 1 cagcgggtgtgagcaggaacg.....gtcattttttttattg 249

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	249	100.0	867	22	AAF91875	Human secreted pro
2	249	100.0	5510	22	AAH73337	Human cervical can
3	36.2	14.5	7784	24	ABL34424	Human immune syste
4	36	14.5	2529	18	AAH7468	Hamster Ubiquitin/
5	35.4	14.2	151826	21	AAH22291	BAC containing rep
6	35.2	14.1	619	21	AAH80495	Human colon cancer
7	34.8	14.0	1734	22	AAH17416	Human CDNA sequenc
8	34.8	14.0	2364	21	AAH70246	Plasmodium falcipar
9	34.4	13.8	15416	24	ABL34231	Human immune syste

10	34.4	13.8	15416	24	AAH61453	Human gene regulat
11	34.2	13.7	7784	24	ABL34425	Human immune syste
12	34	13.7	1900	20	AAH25197	Nucleotide sequenc
13	34	13.7	6298	22	AAH45358	Chemically pretrea
14	33.8	13.6	540	22	AAH10129	Human CDNA clone (
15	33.8	13.6	5536	24	ABL32179	Human immune syste
16	33.8	13.6	7784	24	ABL34424	Human immune syste
17	33.6	13.5	13104	19	AAH52167	Streptococcus pneu
18	33.6	13.5	110000	22	AAH84800	Nucleotide sequenc
19	33.4	13.4	1267	22	AAH21012	Bovine-derived DNA
20	33.4	13.4	6306	22	AAH85746	Human immune/haema
21	33.4	13.4	8395	20	AAH13154	Enterococcus faeca
22	33	13.3	881	17	AAH40048	Human tumour necro
23	33	13.3	881	21	AAH28149	Human TR2 receptor
24	33	13.3	1704	19	AAH34509	Human TNF receptor
25	33	13.3	1704	20	AAH83763	Human tumour necro
26	32.8	13.2	265	17	AAH28133	Senescence-related
27	32.8	13.2	8032	24	AAH61408	Human gene regulat
28	32.6	13.1	5914	24	ABL34169	Human immune syste
29	32.6	13.1	6398	23	ABL13526	Drosophila melanog
30	32.4	13.0	5062	22	AAH46966	Tumour suppressor
31	32.2	12.9	500	21	AAH95256	Cat flea head and
32	32.2	12.9	1804	23	AAH83281	DNA encoding novel
33	32.2	12.9	3069	22	AAH54988	S. epidermidis gen
34	32.2	12.9	3206	22	AAH55018	S. epidermidis gen
35	32.2	12.9	4249	22	AAH54769	S. epidermidis gen
36	32.2	12.9	5536	24	ABL32178	Human immune syste
37	32.2	12.9	6351	24	ABL33745	Human immune syste
38	32.2	12.9	6351	24	ABL34587	Human immune syste
39	32	12.9	5536	24	ABL32179	Human metastasis a
40	32	12.9	17341	21	AAH14872	Human immune syste
41	32	12.9	32150	22	AAH34454	Genomic DNA sequen
42	31.8	12.8	407	22	ABA20983	Human nervous syst
43	31.8	12.8	407	22	AAH89467	Human digestive sy
44	31.8	12.8	1383	24	AAH95277	Long terminal repe
45	31.8	12.8	5314	24	ABL32160	Human immune syste

#### ALIGNMENTS

RESULT 1	
AAF91875	AAF91875 standard; CDNA: 867 BP.
XX	XX
AC	AAF91875;
XX	XX
DT	22-MAY-2001 (first entry)
XX	XX
DE	Human secreted protein-encoding gene 18 CDNA clone HPPFK57, SEQ ID NO:28.
KW	Human; secreted protein; proliferative disorder; cancer; tumour;
KW	focal abnormality; developmental abnormality; haematopoietic disorder;
KW	immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;
KW	inflammation; allergy; neurological disorder; Alzheimer's disease;
KW	Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW	cardiovascular disorder; angiogenic disorder; kidney disorder;
KW	gastrointestinal disorder; pregnancy-related disorder;
KW	endocrine disorder; infection; wound healing; vulnery;
KW	cell culture; chemotaxis; food additive;
KW	binding partner identification; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO200118022-A1.
XX	XX
PD	15-MAR-2001.
XX	XX
PF	31-AUG-2000; 2000WO-US24008.
XX	XX
PR	03-SEP-1999; 99US-0152315.
PR	03-SEP-1999; 99US-0152317.









Db 412 CAGTAATTCGGCTTTAGATGCATCTCTAGCAAGTACTAGCTAAAGCTGCACATTACTTCCTC 413  
 Oy 199 taaaaacagaaaaaagacatcgatgatatgcatattttttt 243  
 1 111 11 11111 111 1 11111 111  
 Db 412 TTAACCAACAGAAAAACGTATATTGCCAGATGTAAGTTGTTT 368  
 RESULT 7  
 ID AAH17416  
 XX AAH17416 standard; cDNA; 1734 BP.  
 XX  
 AC AAH17416;  
 CC  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human CDNA sequence SEQ ID NO:16860.  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss  
 OS Homo sapiens.  
 XX  
 XX Homo sapiens.  
 XX  
 XX EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J.  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS  
 PS Claim 8; SEQ ID 16860; 2537PP + CD ROM; English.  
 CC  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 SQ Sequence 1734 BP; 371 A; 484 C; 428 G; 451 T; 0 other;

Query Match	14.0%;	Score 34.8;	DB 22;	Length 1734;
Best Local Similarity	53.7%;	Pred. No. 1.5;		
Matches	72;	Conservative	0;	Mismatches 62;
			Indels	0;
			Gaps	0.
QY 115	accatgctggggtgctctaaatgctctggaaccccgatggtgtaaaaaagctggagtgaaa	174		
DB 1293	accctgattggnactcatgactgactggaataactacgtaaaccaagtgtcgaagaag	1352		
QY 175	gctcagcatcacatgtattactttaaaaaacagaaaaaagacatgatatgtaatgctca	234		
DB 1353	ccagagctctagtgattagttaacttaactcagactcgtatattaaagatctctctttttct	1412		
QY 235	ttttttttttatlg	248		
DB 1413	tttttttttttttg	1426		

RESULT	8
AAAT70246/c	
ID	AAAT70246 standard; DNA; 2364 BP.
XX	
AC	AAAT70246;
XX	
DT	07-NOV-2000 (first entry)
DE	Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:379.
XX	
XX	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW	antimalarial; malaria; protozoicide; infection; insecticide; ds.
OS	Plasmodium falciparum.
XX	
PN	WO200025728-A2.
XX	
PD	11-MAY-2000.
XX	
PF	05-NOV-1999; 99WO-0526796.
XX	
PR	05-NOV-1998; 98US-0107131.
XX	
PA	(HOEF/) HOFFMAN S.
PA	(CARU/) CARUCCI D.
PA	(GARD/) GARDNER M.
PA	(VENT/) VENTER J C.
XX	
PI	Hoffman S, Carucci D, Gardner M, Venter JC;
XX	
DR	WPI: 2000-365347/31.
XX	
PT	Proteins encoded by chromosome 2 of the human malarial parasite,
PT	Plasmodium falciparum, useful as antimalarial vaccines and in the
PT	diagnosis of P.falciparum infection -
XX	
PS	Disclosure: Page 558-559; 577pp; English.
XX	
CC	The present invention describes proteins and their fragments (I) encoded
CC	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC	Also described are: (I) nucleotide sequences; (II) encoding (I); and (2)
CC	vaccines against P. falciparum infection comprising (I) or (II).
CC	(I) and (II) are useful for the development of vaccines against
CC	P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC	antibody raised to immunogens comprising the sequences of (I), are
CC	useful in the detection of infection with P. falciparum. Furthermore,
CC	(I) (especially when they are rifins or secreted or membrane proteins)
CC	can aid the identification of drugs to treat or prevent P. falciparum
CC	infection, or they can be used to identify drug resistance in
CC	P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC	subsequent identification of proteins encoded by it will help to expand
CC	our understanding of parasite biology, a process hampered by the
CC	complexity of the parasitic lifecycle, and provide new targets for
CC	vaccine and drug development. Parasite resistance to drugs and mosquito
CC	resistance to insecticides have led to a resurgence of malaria in many
CC	parts of the world, and there is a pressing need for vaccines and new





DB 1891 TGTAAATATGATGATTTCAAAATGATTTATATACGCTTGATTAATAATTAAACAG 1832  
 QY 170 tgaagctcagcatalacatgtattacttcaaaacagaaagacatgtagat 229  
 DB 1831 AATAGTACAGGTTTGTGTCATTATTAATAAACAGCATTTGAAATAAACATATTTT 1772  
 QY 230 gtctatttttttttatt 247  
 DB 1771 CATTAATGCTGTTTTTATT 1754

## RESULT 13

AA545358  
 ID AA545358 standard; DNA; 6298 BP.

AC AA545358;

DT 18-DEC-2001 (first entry)

DE Chemically pretreated genomic DNA associated with cell cycle #32.

XX Cell cycle; human; Cpg dinucleotide; cytosine methylation; HIV; aging;  
 KM human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
 KM graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
 KM arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
 KM immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
 KM PCR primer.

OS Homo sapiens.

PN WO200168911-A2.

PD 20-SEP-2001.

PF 15-MAR-2001; 2001WO-EP02945.

PR 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPITG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2001-602751/68.

PT Designing primers and probes for analysing diseases associated with  
 cytosine methylation state e.g. arthritis, cancer, aging,  
 arteriosclerosis comprising fragments of chemically modified genes  
 associated with cell cycle -

PS Claim 1; SEQ ID NO 63; 28pp; English.

XX Sequences AA545296-AA545520 represent chemically pretreated genomic DNA  
 molecules associated with the cell cycle and specific PCR primers of the  
 invention. The sequences are useful for detecting the methylation state  
 of all Cpg dinucleotides in a sequence and therefore for analysing  
 associated diseases. By analysing cytosine methylations in the pretreated  
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
 of existing diseases or the predisposition to specific diseases can be  
 ascertained. The parameters may be compared to another set of genetic  
 and/or epigenetic parameters, the differences serving as basis for  
 CC diagnosis and/or prognosis events which are disadvantageous to patients.  
 CC The sequences of the invention are useful for the diagnosis and therapy  
 of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
 CC aging, glomerular disease, Lewy body disease, arthritis,  
 CC arteriosclerosis, solid tumours and cancers.

XX Sequence 6298 BP; 1583 A; 182 C; 1402 G; 3131 T; 0 other;

Query Match 13.7%; Score 34; DB 22; Length 6298;  
 Best Local Similarity 56.1%; Pred. No. 4.3;  
 Matches 64; Conservativity 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 134 agtgcctgtgaatccgatgtggaagcgtgaggtggaagcctcagcatgatt 193  
 DB 3416 agtgcctgtgaatccgatgtggaagcgtgaggtggaagcctcagcatgatt 3475  
 QY 194 tacttaaaacagaaagacatgtatgatagtcatttttttttttatt 247  
 DB 3476 tttctataataaaagaaatattgataattattttttttttttttt 3529

## RESULT 14

AAH10129/C  
 ID AAH10129 standard; cDNA; 540 BP.

AC AAH10129;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO:6964.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 full-length cDNAs defined in the specification, and for the detection  
 and/or diagnosis of the abnormality of the proteins encoded by the  
 full-length cDNAs -

PS Claim 3; SEQ ID 6964; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 09:14:55 ; Search time 38.87 Seconds  
(Without alignments)  
1573.521 Million cell updates/sec

Title: US-09-757-781-21

Sequence: 1 cagcgggtggtgcaggaagc.....gtctattttttttatg 249

Scoring table: IDENTITY-NUC  
Gap 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*

2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*

3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*

4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*

5: /cgn2\_6/ptodata/2/1na/PCtus.COMB.seq:\*

6: /cgn2\_6/ptodata/2/1na/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	14.5	2529	US-09-051-969A-5	Sequence 5, Appl
2	35.2	14.1	619	US-09-328-111-579	Sequence 579, App
3	31.4	12.6	579	US-09-146-950-3	Sequence 3, Appl
4	31.4	12.6	591	US-09-146-950-19	Sequence 19, Appl
5	31.4	12.6	1596	US-09-146-950-17	Sequence 17, Appl
6	31.4	12.6	1724	US-08-509-024-1	Sequence 1, Appl
7	31.4	12.6	1724	US-09-333-279-1	Sequence 1, Appl
8	31.4	12.6	1929	US-09-146-950-1	Sequence 1, Appl
9	31.4	12.6	4622	US-08-509-024-6	Sequence 6, Appl
10	31.4	12.6	4622	US-09-333-279-6	Sequence 6, Appl
11	30.4	12.2	2646	US-08-539-304A-5	Sequence 5, Appl
12	30.4	12.2	2663	US-08-136-743B-3	Sequence 3, Appl
13	30.4	12.2	3252	US-09-118-442-1	Sequence 1, Appl
14	30.4	12.2	3252	US-09-677-064-1	Sequence 1, Appl
15	29.8	12.0	1724	US-07-847-010-18	Sequence 18, Appl
16	29.6	11.9	860	US-07-847-010-18	Sequence 18, Appl
17	29.6	11.9	2658	US-08-773-608A-1	Sequence 1, Appl
18	29.4	11.8	524	PCT-US95-08295-22	Sequence 22, Appl
19	29.4	11.8	4875	US-08-460-739-1	Sequence 1, Appl
20	29.2	11.7	19124	US-08-487-826B-13	Sequence 13, Appl
21	29	11.6	279	US-08-686-878A-50	Sequence 50, Appl
22	29	11.6	279	US-08-721-489-4	Sequence 4, Appl
23	28.8	11.6	684	US-08-577-463A-1	Sequence 3, Appl
24	28.8	11.6	2946	US-09-175-928-3	Sequence 3, Appl
25	28.4	11.4	2622	US-08-766-014-23	Sequence 23, Appl
26	28.4	11.4	2902	US-08-714-918-95	Sequence 95, Appl
27	28.4	11.4	2902	US-09-265-315-95	Sequence 95, Appl

c 28	28.4	11.4	2902	4	US-09-265-315-95	Sequence 95, Appl
c 29	28.4	11.4	2902	4	US-09-265-417-95	Sequence 95, Appl
c 30	28.4	11.4	5300	1	US-08-766-014-1	Sequence 1, Appl
c 31	28.2	11.3	618	4	US-08-961-810-17	Sequence 17, Appl
c 32	28.2	11.3	618	4	US-08-352-903D-17	Sequence 17, Appl
c 33	28.2	11.3	2075	1	US-08-238-163-3	Sequence 3, Appl
c 34	28	11.2	392	1	US-08-469-421-8	Sequence 8, Appl
c 35	28	11.2	392	1	US-08-250-975-8	Sequence 8, Appl
c 36	28	11.2	392	2	US-08-605-002A-8	Sequence 8, Appl
c 37	28	11.2	392	2	US-08-950-449A-8	Sequence 8, Appl
c 38	28	11.2	392	5	PCT-US94-10529-8	Sequence 8, Appl
c 39	28	11.2	2612	4	US-09-105-390-7	Sequence 7, Appl
c 40	28	11.2	6124	4	US-08-213-419B-3	Sequence 3, Appl
c 41	27.8	11.2	886	1	US-08-469-427A-1	Sequence 1, Appl
c 42	27.8	11.2	886	2	US-08-609-443B-1	Sequence 1, Appl
c 43	27.8	11.2	886	2	US-08-569-063C-1	Sequence 1, Appl
c 44	27.8	11.2	1002	4	US-08-960-780-43	Sequence 43, Appl
c 45	27.8	11.2	1002	4	US-09-073-898-43	Sequence 43, Appl

## ALIGNMENTS

RESULT 1  
US-09-051-969A-5  
Sequence 5, Application US/09051969A  
Patent No. 6063598  
GENERAL INFORMATION:  
APPLICANT: ENEKEL, BARBARA  
APPLICANT: CANNON, FRANK  
APPLICANT: BERGMANN, KLAUS  
APPLICANT: NOE, WOLFGANG  
TITLE OF INVENTION: INTENSIVE HOMOLOGOUS PROMOTER OBTAINED  
TITLE OF INVENTION: FROM HAMSTERS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/051,969A  
FILING DATE: 1998-09-30  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FLESHNER, RAZ E.  
REGISTRATION NUMBER: 34,331  
REFERENCE/DOCKET NUMBER: 0652.1690000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2529 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2418..2465  
US-09-051-969A-5  
Query Match 14.5%; Score 36; DB 3; Length 2529;  
Best Local Similarity 50.9%; Pred. No. 0.042;





```

; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-146-950-17

```

```

Query Match      12.6%; Score 31.4; DB 4; Length 1596;
Best Local Similarity 51.0%; Pred. No. 1;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

Qy 53 gggctcggggatctcggtggtcccatctccttcattgttctgaacatccgtatgt 112
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 587 GGGTGTCTGACTCTGCGTCCCTCTTGCACCCCTTGCCCGGGCTGGAGTGGCGT 528
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Qy 113 aaacatggtcgtggtgtaaaagtgcctgtgaatcccatgtgtaaaagtgaagtg 172
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 527 AAGCGGGGACAGCGGGCGAGTGGTCCCGCTGACAGATGAGAAGTGGCTGGGCTGC 468
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Qy 173 aagtcagcataccatgtattact 197
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 467 AGCCACACACGCGCTTCTGTCT 443
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||

```

```

RESULT 6
US-08-509-024-1/c
; Sequence 1, Application US/08509024B
; Patent No. 6291207
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/08/509,024B
; CURRENT FILING DATE: 1995-07-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-509-024-1

```

```

Query Match      12.6%; Score 31.4; DB 4; Length 1724;
Best Local Similarity 51.0%; Pred. No. 1.1;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

Qy 53 gggctcggggatctcggtggtcccatctccttcattgttctgaacatccgtatgt 112
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 774 GGGTGTCTGACTCTGCGTCCCTCTTGCACCCCTTGCCCGGGCTGGAGTGGCGT 715
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Qy 113 aaacatggtcgtggtgtaaaagtgcctgtgaatcccatgtgtaaaagtgaagtg 172
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 714 AAGCGGGGACAGCGGGCGAGTGGTCCCGCTGACAGATGAGAAGTGGCTGGGCTGC 655
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Qy 173 aagtcagcataccatgtattact 197
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 654 AGCCACACACGCGCTTCTGTCT 630
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||

```

```

RESULT 7
US-09-333-279-1/c

```

```

; Sequence 1, Application US/09333279
; Patent No. 6303336
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/09/333,279
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-333-279-1

```

```

Query Match      12.6%; Score 31.4; DB 4; Length 1724;
Best Local Similarity 51.0%; Pred. No. 1.1;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

Qy 53 gggctcggggatctcggtggtcccatctccttcattgttctgaacatccgtatgt 112
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 774 GGGTGTCTGACTCTGCGTCCCTCTTGCACCCCTTGCCCGGGCTGGAGTGGCGT 715
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Qy 113 aaacatggtcgtggtgtaaaagtgcctgtgaatcccatgtgtaaaagtgaagtg 172
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 714 AAGCGGGGACAGCGGGCGAGTGGTCCCGCTGACAGATGAGAAGTGGCTGGGCTGC 655
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Qy 173 aagtcagcataccatgtattact 197
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 654 AGCCACACACGCGCTTCTGTCT 630
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||

```

```

RESULT 8
US-09-146-950-1/c
; Sequence 1, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1929
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (297)...(875)
; US-09-146-950-1

```

```

Query Match      12.6%; Score 31.4; DB 4; Length 1929;
Best Local Similarity 51.0%; Pred. No. 1.1;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

Qy 53 gggctcggggatctcggtggtcccatctccttcattgttctgaacatccgtatgt 112
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 777 GGGTGTCTGACTCTGCGTCCCTCTTGCACCCCTTGCCCGGGCTGGAGTGGCGT 718
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Qy 113 aaacatggtcgtggtgtaaaagtgcctgtgaatcccatgtgtaaaagtgaagtg 172
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 717 AAGCGGGGACAGCGGGCGAGTGGTCCCGCTGACAGATGAGAAGTGGCTGGGCTGC 658
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Qy 173 aagtcagcataccatgtattact 197
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 657 AGCCACACACGCGCTTCTGTCT 633
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||

```



```

; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The University of Pennsylvania
; STREET: Suite 330
; STREET: 3700 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19104-3246
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,743B
; FILING DATE: 10/14/93
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 3957-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5459063e
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2663 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-136-743B-3

Query Match      12.2%; Score 30.4; DB 1; Length 2663;
Best Local Similarity 63.9%; Pred. No. 2.8;
Matches 46; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 176 ctacagatccatgattcttcttaaaacagaaaaagacatgctgatatgtctat 235
Db 2301 cttatataaaatgacatcactatataaataatataatataatataatataat 2242
Qy 236 ttttttttatt 247
Db 2241 atttttttttt 2230

RESULT 13
US-09-118-442-1/c
; Sequence 1, Application US/09118442B
; Patent No. 6197561
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; APPLICANT: Wang, Xun
; APPLICANT: Bowen, Benjamin A.
; TITLE OF INVENTION: Genes Controlling Phylate Metabolism in
; FILE REFERENCE: 0706
; CURRENT APPLICATION NUMBER: US/09/118,442B
; EARLIER FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/055,446
; EARLIER FILING DATE: 1997-08-11
; EARLIER APPLICATION NUMBER: 60/055,526
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: 60/053,944
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1

```

```

; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (258)...(2666)
; NAME/KEY: misc_feature
; LOCATION: (1)...(3252)
; OTHER INFORMATION: n = A,T,C or G
; US-09-118-442-1

Query Match      12.2%; Score 30.4; DB 4; Length 3252;
Best Local Similarity 50.0%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 92 gtctgacatccctgattgtaaaccaatgctgggtgctaaagtgcctgtgaatccga 151
Db 1862 gttcaaggcactcttaatatagacgattcttgagcgacacactcctgttagagaaagct 1803
Qy 152 tgtgaaaagctggaagtgaaagctcagcataccatglatlactttaaaacagaaaa 211
Db 1802 tgcgtcaaatgctgacgaagaattgctcattcaaaagtctgaagctcgaagaaactcctga 1743
Qy 212 aaagacatgctatgatatgtctattttttt 243
Db 1742 taatagctgcctcaatttttcaattttctttt 1711

RESULT 14
US-09-677-064-1/c
; Sequence 1, Application US/09677064
; Patent No. 6291224
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; TITLE OF INVENTION: Genes Controlling Phylate Metabolism in
; FILE REFERENCE: 0706D
; CURRENT APPLICATION NUMBER: US/09/677,064
; EARLIER FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; PRIOR FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (258)...(2666)
; NAME/KEY: misc_feature
; LOCATION: (1)...(3252)
; OTHER INFORMATION: n = A,T,C or G
; US-09-677-064-1

Query Match      12.2%; Score 30.4; DB 4; Length 3252;
Best Local Similarity 50.0%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 92 gtctgacatccctgattgtaaaccaatgctgggtgctaaagtgcctgtgaatccga 151
Db 1862 gttcaaggcactcttaatatagacgattcttgagcgacacactcctgttagagaaagct 1803

```

QY 152 tctggaagaagctggaggtgaaagctcagcatatcatttactttaaacagaaaa 211  
DB 1802 TGGTGGCAATAGGTGAAGAAATGGCTCATCAAACTTTGTAAGCTCACTGAAAACTCCTGA 1743  
QY 212 aaagacatgtatgatatgtctatctttttt 243  
DB 1742 TAATAGCTGCCTCAATTTTCAATTTTCTTTT 1711

## RESULT 15

PCT-US96-12374-1/c  
; Sequence 1, Application PC/RUS9612374  
; GENERAL INFORMATION:  
; APPLICANT: Northwestern University  
; TITLE OF INVENTION: Herpes Virus Entry Mediator  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dressler, Goldsmith, Milnamow & Katz, Ltd.  
; STREET: 180 N. Stetson, Suite 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PC-DOS/MS-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/12374  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Northrup, Thomas E.  
; REGISTRATION NUMBER: 33,268  
; REFERENCE/DOCKET NUMBER: NOR446P020PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 616-5400  
; TELEFAX: (312) 616-5460  
; TELEX: --  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1724 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 294..1145  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 294..1142  
; PCT-US96-12374-1

Query Match 12.0%; Score 29.8; DB 5; Length 1724;

Best Local Similarity 50.3%; Pred. No. 3.5;

Matches 73; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 53 gggctcgggagatcgtggtccatcttcctcatgttctgaacatccctgtattgt 112  
DB 774 GGGTGTCTGACTGTGCTGCTCCCTTCGACCTGTGCGCCGGGCTGAGGTGCGCT 715  
QY 113 aaacacatgtcgtggtgctaaagtctgtaatcccgatgtgaaaaagctgaggtga 172  
DB 714 AACGGCGGACAGCGGCGCAGTGGTCCCGTCTGAGAGATGACAGAGTGGCTGGCTGC 655  
QY 173 aagctcagcatcacatgtattact 197  
DB 654 AGCCACACACAGCGGTCTCTGTCT 630

Search completed: September 21, 2002, 10:17:28  
Job time: 3753 sec



(Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT

174 a 126 c 114 g 122 t

Query Match 100.0%; Score 249; DB 9; Length 536;  
Best Local Similarity 100.0%; Pred. No. 1.1e-51;  
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagcgggtggtgaggaagcctcctcgcgtcagatagagtggtggtctcg 60  
|||||  
DB 497 CACGGGTGTTGAGGAGAGCTCACTCGGCTAGTATAGAGTGTGTGGTCTCG 438  
|||||  
OY 61 gggatctggtggtcccatctcctcattgttcgaaacccgtattgtaaccatg 120  
|||||  
DB 437 GGGATCTCGGTGCTCCCATCTTCTCATTTGTGAAACATCCTGATTTAAACATG 378  
|||||  
OY 121 gctggggtgctaaagtgcctgtaatccgagtgtgaaagcgtggaagctcag 180  
|||||  
DB 377 GCGGGGTGCTAAAGTGCCTGTGAATCCGAGTGTGAAAGCTGAGAGCTCAG 318  
|||||  
OY 181 catcacatgtattactttaaaacagaaaaagacatgtatgatatgtctatttt 240  
|||||  
DB 317 CATACCATGTATTACTTTTAAAAACAGAAAAAGACATGATGATGATGCTATTTTT 258  
|||||  
OY 241 ttattatgg 249  
|||||  
DB 257 TTTTATTGG 249

RESULT 2

LOCUS A1497808 536 bp mRNA linear EST 14-APR-1999

DEFINITION tm89f02.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2165307 3',  
mRNA sequence.

ACCESSION A1497808

VERSION A1497808

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 536)

AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAAP), Tumor Gene Index

JOURNAL Unpublished (1998)

COMMENT Contact: Robert Strausberg, Ph.D.

Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Bonaldo, Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.lnl.gov/bbrp/image/image.html

Insert Length: 779 Std Error: 0.00

Seq primer: -400P from Gldco

High quality sequence stop: 468.

Location/Qualifiers

1. 536

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2165307"

/clone\_lib="NCI\_CGAP\_Brn25"

/tissue\_type="anaplastic oligodendroglioma"

/lab\_host="DH10B"

/note="Organ: brain; Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTACCAATCTGAAGTGGAGCGGCGCCAGATGGATTTTATTTTATTTT T 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

174 a 126 c 114 g 122 t

Query Match 100.0%; Score 249; DB 9; Length 536;  
Best Local Similarity 100.0%; Pred. No. 1.1e-51;  
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagcgggtggtgaggaagcctcctcgcgtcagatagagtggtggtctcg 60  
|||||  
DB 497 CACGGGTGTTGAGGAGAGCTCACTCGGCTAGTATAGAGTGTGTGGTCTCG 438  
|||||  
OY 61 gggatctggtggtcccatctcctcattgttcgaaacccgtattgtaaccatg 120  
|||||  
DB 437 GGGATCTCGGTGCTCCCATCTTCTCATTTGTGAAACATCCTGATTTAAACATG 378  
|||||  
OY 121 gctggggtgctaaagtgcctgtaatccgagtgtgaaagcgtggaagctcag 180  
|||||  
DB 377 GCTGGGTGCTAAAGTGCCTGTGAATCCGAGTGTGAAAGCTGAGAGCTCAG 318  
|||||  
OY 181 catcacatgtattactttaaaacagaaaaagacatgtatgatatgtctatttt 240  
|||||  
DB 317 CATACCATGTATTACTTTTAAAAACAGAAAAAGACATGATGATGATGCTATTTTT 258  
|||||  
OY 241 ttattatgg 249  
|||||  
DB 257 TTTTATTGG 249

RESULT 3

LOCUS A1038061 561 bp mRNA linear EST 28-AUG-1998

DEFINITION ox21b05.x1 Soares fetal\_liver\_spleen\_INFIS\_S1 Homo sapiens cDNA

clone IMAGE:1656945 3', mRNA sequence.

ACCESSION A1038061

VERSION A1038061

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 561)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1466 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 465.

Location/Qualifiers

1. 561

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1656945"

/clone\_lib="Soares\_fetal\_liver\_spleen\_INFIS\_S1"

/sex="male"

/dev\_stage="20 week post conception fetus"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)

with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;

```

FEATURES
source
location/Qualifiers
1..626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4295392"
/clone_1lb="NH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (clontech);

```

Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPES/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=st2-RC3-H00585-160>)  
300-022-d048t3=2000-03-16&t4=1)  
Seq primer: puc 18 forward



High quality sequence start: 20  
High quality sequence stop: 617.

FEATURES  
source  
1. 618  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="H0585"  
/dev\_stage="Adult"  
/note="Organ: head, neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT  
178 a 146 c 124 g 170 t

ORIGIN

Query Match 97.4%; Score 242.6; DB 9; Length 618;  
Best Local Similarity 98.4%; Pred. No. 4.2e-50;

Matches 245; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cagcggctggtgaggaagcctcactcctcgcgtcagtagatgagtggtggtctcg 60  
|||||  
Db 303 CACCGGTGGGCGAGAGACTCACTCGCGTCACTATTAGAGTGTGTGTGGTCTCG 244  
|||||  
QY 61 gggatctcgtggtcctccatctcctcattgttctgaaacatcctgattgtaacatg 120  
|||||  
Db 243 GGGATCTCGGTGCTCCCACTTCCTTCATTGTTGACATCTCTGATTGTAGACCATG 184  
|||||  
QY 121 gctgggtgctgaagtgccctgtaatccgatgtggaagagctggagtggaagctcag 180  
|||||  
Db 183 GCGGGGTGCTTAAGTCCCTGTGAATCCGATGGGAGAAAGCTGGAGGCTCAG 124  
|||||  
QY 181 cataccatgtatttactttaaaacagaaaaaagacatgatatgtctattttt 240  
|||||  
Db 123 CATACCATGATTACTTTAAATACAGAAAAAAGACATGATGATGCTATTTT 64  
|||||  
QY 241 ttattatgg 249  
|||||  
Db 63 TTTTACTCG 55

RESULT 6

LOCUS AW242451 540 bp mRNA EST 14-DEC-1999

DEFINITION xm99a04.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2692302 3',

ACCESSION AW242451  
VERSION AW242451.1 GI:6576205

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 540)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.

Emmert-Buck, M.D., Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco

High quality sequence stop: 463.

FEATURES  
source  
1. 540  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2692302"  
/clone\_id="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP hybridization, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bernaldo."

BASE COUNT  
175 a 126 c 113 g 126 t

ORIGIN

Query Match 96.8%; Score 241; DB 9; Length 540;  
Best Local Similarity 98.0%; Pred. No. 1.1e-49;

Matches 244; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 cagcggctggtgaggaagcctcactcctcgcgtcagtagatgagtggtggtctcg 60  
|||||  
Db 501 CACCGGTGGGCGAGAGACTCACTCGCGTCACTATTAGAGTGTGTGTGGTCTCG 442  
|||||  
QY 61 gggatctcgtggtcctccatctcctcattgttctgaaacatcctgattgtaacatg 120  
|||||  
Db 441 GGGATCTCGGTGCTCCCACTTCCTTCATTGTTGACATCTCTGATTGTAGACCATG 382  
|||||  
QY 121 gctgggtgctgaagtgccctgtaatccgatgtggaagagctggagtggaagctcag 180  
|||||  
Db 381 GTTGGGTGTTAAAGTCCCTGTGAATCCCGATGGGAGAAAGCTGGAGGCTCAG 322  
|||||  
QY 181 cataccatgtatttactttaaaacagaaaaaagacatgatatgtctattttt 240  
|||||  
Db 321 CATACCATGATTACTTTAAATACAGAAAAAAGACATGATGATGCTATTTT 262  
|||||  
QY 241 ttattatgg 249  
|||||  
Db 261 TTTTATTGG 253

RESULT 7

LOCUS AI333325 486 bp mRNA EST 13-FEB-1999

DEFINITION q905D01.x1 Soares\_NHMPU\_S1 Homo sapiens cDNA clone IMAGE:1931593

ACCESSION AI333325  
VERSION AI333325.1 GI:4069884

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 486)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Insert length: 782 Std Error: 0.00  
Seq primer: -40UP from Gibco

High quality sequence stop: 447.

Location/Qualifiers  
1. 486  
/organism="Homo sapiens"

```

/adb_xref="taxon:9606"
/clone_image="1931593"
/clone_lib="Scares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below): Vector: pRTT3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NBHMU, and fetal heart NBH19W) were mixed, and ss circles
was made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265523,
340488-345479, and 484488-489479."

```

Query Match	95.28;	Score 237;	DB 9;	Length 486;
Best Local Similarity	100.0%;	Pred. No. 1.1e-48;		
Matches 237;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 13 cagaagagtcacatccgcgcgtcagcattcagagcgtgtgtcgttcgcggagatcccgagc 72  
Db 486 CAGGAAGCTCACTCGGGTCAGTATTGAGTGTGTGTGGGTCTCGGGGATCTCGGTG 427  
QY 73 gtcgccatctccctcattcattgttcgaaacatccgtattgttaaacaatgctcggagtccta 132  
Db 426 GCTCCACATCTTCCTTCATCTGTTCTCAAAATCCTGTATTGTATAACCATGCGTGGGGGCTA 367  
QY 133 aatgcctcgtgaatcccgatgtgaaaaagctcgaggggtgaaagctcagatccatcattat 192  
Db 366 AAGTGCCCTGTGATCCCATGTGGAAAAAGCTGGAGGTGAAGCTCAGCATGTCCATGTAT 307  
QY 193 ttaactttaaaaaacagaaaaaaagacatgtatgtatgtcctattctttttattatcg 249  
Db 306 TTACTTTAAAAACAAAAAAGACATGTATGTGATATGCTATTTTTTTTTTTTATATGG 250

RESULT	8
A1952086/c	
LOCUS	486 bp mRNA linear EST_06-SEP-1996 wv46glll.x1 NCI-CGAP_Lu28 Homo sapiens cDNA clone IMAGE:2546756'
DEFINITION	mRNA sequence.
ACCESSION	A1952086
VERSION	A1952086.1 GI:5744396
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D.
JOURNAL	
COMMENT	

FEATURES  
source  
Location/Qualifiers  
1. .486  
High quality sequence stop: 413.  
Information can be found through the I. M. A. G. E. Consortium/TLNM at: [www-bio.1lnl.gov/bdrr/image/image.html](http://www-bio.1lnl.gov/bdrr/image/image.html)  
Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christina Frange, The I. M. A. G. E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2546756"
/clone_1fb="NCI-CGAP_Lu28"
/tissue_type="Two pooled squamous cell carcinomas"
/lab_host="DH10B"
/site="Organ: lung; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT library constructed by Life Technologies."
BASE COUNT      164 a      112 c      100 g      110 t
ORIGIN

```

Query Match	95.2%	Score 237	DB 9	Length 486
Best Local Similarity	100.0%	Pred. No. 1.1e-48		
Matches 237	Conservative 0	Mismatches 0	Indels 0	Gaps 0

[illegible][illegible]

```

FEATURES
    source
        1..545
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1713028"
            /clone_11b="Soares_pregnant_uterus_NBDPV"
            /sex="female"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
            Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
            oligo(dT) primer [5',
            AACTGGAGAAATTCGCGGCCCTTTTGTGTGTGTGTGT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Not I and cloned into the Not I

```

BASE COUNT 172 a 128 c 116 g 129 t  
ORIGIN

Query Match	93.7%	Score 233.2	DB 9	Length 545
Best Local Similarity	98.4%	Pred. No. 9.3e+48		
Matches 246; Conservative	0;	Mismatches 3;	Indels 1;	Gaps 1

QY	1	cagcgctggctgcgaagaagcctcaacctgcgcacgacatlaaggtgtgtgtgtgggctcgg	60
Db	497	CAGCGGTGGTGGCAGGAAGCTACCTCCCGCAGCATATTAGAGTGTGTGTGGGCTCTGG	438
QY	61	gggacatcgtgtgtgtgtcccatcttcctt-cattgtcttgacaactcctgtatgtglaaacat	119
Db	437	GGGATCTCGGTGGCGCCCATCTTCCTTCATGTGTTGGACATCCGATTTGTAACCAT	378
QY	120	ggcctggaggtgtcaaaagtgcctgtgaatcccgatgtgtgaaaaagcctggaggtgaaaaagctca	179
Db	377	GGCTGGGGTGTCTAAAGTCCCTGTGAATCCCATGTGTGAAAAAGCTGTGGAGGTGTAAGCTCA	318
QY	180	gcataccatgtatctacttlaaaaaacagaaaaaagacatgtatgtatgtctatlltt	239
Db	317	GCATACCATGTATTTACTTTAAAAACAGAAAAAAGACATGTATGATATGCTATTTT	258
QY	240	tttttatatgg 249	
Db	257	TTTTTATTGG 248	

	RESULT	10				
		BE176633/c				
LOCUS	BE176633		687 bp	mRNA	linear	EST 21-JUN-2000
DEFINITION	RC3-HR0585-010400-013-b05 HR0585			Homo sapiens CDNA,		mRNA sequence.
ACCESSION	BE176633					
VERSION	BE176633.1			GI:8639362		
KEYWORDS	EST.					
SOURCE	human.					

```

/clone_lib="HT0585"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      197 a      173 c      143 g      174 t
ORIGIN

```

BASE COUNT	197 a	173 c	143 g	174 t
ORIGIN				
Query Match		93.7%	Score 233.2;	DB: 9;
Best Local Similarity		98.4%	Pred. NO. 8.8e-48;	Length 687;
Matches 246;	Conservative	0;	Mismatches 3;	Indels 1;
				Gaps 1;

QY	1	caagcggatggatggcagaagaagcaccatcccgctcaagatataagatgtggtgtgagtcgg	60
Db	315	CAGCGGATGGATGGCAGGAAGCTCACTCTCGCGCTCAGTATTAAAGTGTGTGGGTCTCG	256
QY	61	ggagatcccgatggtgcctcccatctccctcatatgtgtcttgaacaatcctgtatgttaaacatg	120
Db	255	GGGATTCGCGGGCGCTCCCATCTTCCTTCATGTCTCTGGACATCCGTATTGTAGACCATG	196
QY	121	gctcggagtgctcaaaagtgcctgtgaatcccgatgtggaaaaaagctgggaagttaaagctcag	180
Db	195	GCTCGGGATGCTAAAGTGCCTGTGAAATCCGATGTGGATTAAGCTGGAGGTAAAGCTTAG	136
QY	181	cataccatgtatcttacttactttaaaa-cagaaaaaaagaacatgtatgataatgtctatttt	239
Db	135	CATACCATGTATTACTTTAAAAACAGAAAAAAGACATGTATGATATCTCATATTTT	76
QY	240	tttttttttt 249	
Db	75	TTTTTATTGG 66	

RESULT	11
LOCUS	A1356239/c
DEFINITION	463 bp mRNA linear EST 16-FEB-1999
ACCESSION	A1356239
VERSION	GV65F11.x1 NCI_CGAP_Brn2s Homo sapiens cDNA clone IMAGE:2016909 3'
KEYWORDS	mRNA sequence.
SOURCE	A1356239.1 GI:4107860
ORGANISM	Homo sapiens
TISSUE	EST.
REFERENCE	human.

FEATURES	Location/Qualifiers
source	1. .687
	/organism="Homo sapiens"
	/db_xref="taxon:9606"

FEATURES	Location/Qualifiers
source	1. .463
	/organism="Homo sapiens"

```

BASE COUNT      160 a      107 c      92 g      104 t
ORIGIN
/adb_xref="taxon:9606"
/clone_image="2016909"
/clone_id="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with Not I - oligo(dT) primer [5'
TGTATCCATCATCTGAAGTGGAGGCGCCGCAATAGATTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and Eco RI into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

```

	Query Match	86.3%	Score 215:	DB 9:	Length 463;
	Best Local Similarity	100.0%;	Pred. NO.	3.1e-43;	
	Matches 215:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	35	gataagaagtgtgtgtgtgtgtgtcgcggagatctcgtytgcctccacttcctcaatgtt	94		
Db	463	GTATTAGAGTGTGCTGTGGGCTCGGGGATCTCGTGCCATCTTCCTTAATTGT	404		
OY	95	ctgaacatccgtatttgttaaacatagcgtcggttggtctaagtgccttfgaatcccatgt	154		
Db	403	CTGAACATCCTGATTATTGTAAACCATAGCGCTGGGGTCTTAAAGTCCTGTGTAATCCCGANGT	344		
OY	155	ggaaaaagcttgaaggctgaagagctcacatatcatcttactttaaaacaagaaaaa	214		
Db	343	GCAAAAACCTGGAGGTGAAGAGCTGCAGCATACATGTTATTATTAAAAACGAAAAAAAA	284		
OY	215	gacatgatgatgatgtctatttttttttatgtg	249		
Db	283	GACATGATGATGATATGCTAATTTTTTTTTTATTGG	249		

```

/clone="IMAGE:1913978"
/clone_lib="NCI_CGAP_K1d5"
/tissue_type="2_pooled tumors (clear cell type)"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: pT73d-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cdna was primed with a Not I - oligo(dT) primer [5'
AACGCGAAGATTCGCGGCCCAATATTTTCTTTTCTTTTCTTTTCTTTT
3'],
double-stranded cdna was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

```

[illegible]

FEATURES	SOURCE
<p> <b>LOCUS</b> AI309934/c  <b>DEFINITION</b> 462 bp mRNA linear EST 01-<del>FE</del>B-1999  <b>ACCESSION</b> g071602.x1 NCI-CGAP_Ki65 Homo sapiens CDNA clone IMAGE:191378 3',  <b>VERSION</b> mRNA sequence.  <b>KEYWORDS</b> AI309934 AI309934.1 GI:4004805  <b>SOURCE</b> EST.  <b>ORGANISM</b> human.  <b>REFERENCE</b> Homo sapiens  <b>AUTHORS</b> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  <b>TITLE</b> Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  <b>COMMENT</b> 1 (bases 1 to 462)  NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a>.  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  Tumor Gene Index  Unpublished (1997)  <b>JOURNAL</b> Contact: Robert Strausberg, Ph.D.  Email: cgapsb@remail.nih.gov  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  Emmert-Buck, M.D., Ph.D.  <b>CDNA Library Preparation:</b> M. Bento Soares, Ph.D.  <b>CDNA Library Arrayed by:</b> Greg Lennon, Ph.D.  <b>DNA Sequencing by:</b> Washington University Genome Sequencing Center  <b>Clone distribution:</b> NCI-CGAP clone distribution information can be  found through the I.M.A.G.E.E. Consortium/LNL at:  <a href="http://www-bio.llnl.gov/bdnp/image/image.html">www-bio.llnl.gov/bdnp/image/image.html</a>  <b>Insert Length:</b> 1094 Std Error: 0.00  <b>Seq primer:</b> -40UP from Gidco  <b>High quality sequence stop:</b> 425.  <b>location/Qualifiers</b>  1. 462  /organism="Homo sapiens"  /db_xref="taxon:9606" </p>	

```

RESULT 13
AL369277/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

```



```

/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: ovary; Vector: p773D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCATCTGAAGTGGAGCGCGCGGTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library constructed by Bento Soares and
M.Felima Bonaldo."

```

```

BASE COUNT      147 a      89 c      88 g      103 t
ORIGIN

```

```

Query Match      70.7%: Score 176; DB 9; Length 427;
Best Local Similarity 100.0%; Pred.No. 1.5e-33;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      74 ctcacatcttccttcattgtctgaacacccgtgtattgtaaacatgctgaggctctaa 133
         |||||||
DB      427 CTCCCATCTTCCTTCATTTGTTGGAACATCCTCTATTGTAAACCATGGCTGGGGTCTAA 368
         |||||||
OY      134 agtgccctgtaatcccgatctggaagaagctggaggctgaagctcagcatatgtatt 193
         |||||||
DB      367 ACTGCTGTGAATCCGATGTGGAAGAGCTGAGAGTGAAGCTCAGCATACCATGTATT 308
         |||||||
OY      194 tactttaaaacagaaaaaagacatgtatgtatgtctattttttttatttg 249
         |||||||
DB      307 TACTTTAAAAACGAAAAAAGACATGTATGATATGCTATTTTCTTTTATTGG 252
         |||||||

```

```

Search completed: September 21, 2002, 09:42:10
Job time: 5937 sec

```